Journal of
Experimental Psychology

VOL. 30, NO. 2 FEBRUARY, 1942

BEHAVIOR POTENTIALITY AS A JOINT FUNCTION OF THE AMOUNT OF TRAINING AND THE DEGREE OF HUNGER AT THE TIME OF EXTINCTION [1]

BY C. THEODORE PERIN

Department of Psychology, Institute of Human Relations, Yale University

For the most part, experiments designed to show relations between observed behavior and conditions antecedent to the occurrence of the behavior have been limited to single pairs of variables. In such experiments all known antecedent variables except one are held as constant as experimental conditions will permit. This one variable, sometimes called the experimental variable, is varied systematically by the experimenter, and the concomitant changes in the dependent or consequent variable are noted. Nearly all animal-drive studies have been carried out in the fashion just described.

Crutchfield (1) has pointed out that generalizations are extremely limited when only a single antecedent variable is used, and that investigation of the interaction of antecedent variables in the determination of behavior is impossible when only one antecedent variable is present in any given experiment. Fitts (5) also has recognized the importance, as well as the difficulties, of experiments involving more than one antecedent variable.

The present study is a multi-variable one in that it is designed to show behavior as a mathematical function of two antecedent variables, the degree of training, and the intensity of the hunger drive present at the time the behavior potential is measured. The ap-

[1] This investigation is a part of the coordinated research program of the Institute of Human Relations, Yale University. The writer is greatly indebted to Professor Clark L. Hull, under whose direction the experiment was carried out, for valuable suggestions and criticisms. A number of intricate curve-fitting problems were solved by Mr. Bengt Carlson. The results of an earlier study by Dr. S. B. Williams are reproduced here with his kind permission. The experimental results of this study were contained in a thesis presented by the writer in partial fulfillment of the requirements for the degree of Master of Arts, Yale University, 1940.

proach to the multi-variable problem here employed is distinctly different from the Fisher design which Crutchfield has advocated.[2] Basically the present design is of three-dimensional form with the two experimental variables occupying two of the dimensions, and the measure of behavior potentiality occupying the third dimension.[3] With the data presented in this manner it is possible to express a measure of behavior potentiality as a function of either of the antecedent or experimental variables taken singly (the other remaining constant) or of both antecedent variables acting together.

The term *behavior potentiality* is used here with specific reference to the way in which the behavior is measured. For example, when an albino rat has been trained to press a lever in order to obtain a pellet of food and the hungry animal is then given access to the lever but no food is given when the lever is pressed, the lever-pressing response may be expected to cease after a number of responses have been made. When the non-reinforced responses are counted, the result may be taken as an indication of the response potentiality before the response was extinguished. The potentiality, then, can be measured only after it has ceased to be a potentiality and has become an actuality through occurrence.

Subjects and Apparatus

The subjects for this experiment were 320 male albino rats approximately 90 days old and ranging from 180 to 225 grams in weight.

The apparatus consisted of four modified Skinner boxes of the type used by Ellson (2) and Williams (10). These boxes were partially soundproofed and were ventilated by means of the laboratory suction line. The compartment in which the animal was placed measured one foot along each of its three dimensions. At one end of the cubical compartment was a brass panel and through an aperture in this panel a horizontal bar 55 mm in length could be inserted in the box (Fig. 1). The aperture was automatically covered by a sliding shutter when the bar was withdrawn. During the training period, pressure of fifteen grams on the bar by the rat closed a mercury-cup contact in series with an electromagnetic food release. The resulting action of the magnet released one pellet of food from the food magazine into the food tray. In addition, pressure on the bar activated a magnetic signal marker mounted on a constant-speed, waxed-paper polygraph, thus recording the number of responses and the time required for each response. The pellets were made from a mixture of finely ground calf meal and water. The mixture was molded into a cylindrical shape and after drying it was cut into small cylinders weighing about .05 gram each. The consumption of one pellet thus secured constituted one reinforcement.

[2] Unfortunately, the Fisher procedure yields merely an indication of the degree of dependence of some consequent variable upon a number of antecedent variables, and gives no indication of the nature of the functional relationships involved. For rather gross practical purposes, such as are characteristic of agriculture, the extent of the relationship among variables regardless of their functional nature may be sufficient, but for a science which hopes for precise systematization the determination of joint functional relationships is indispensable. For that reason the Fisher design type of experiment with its characteristic economies of time and effort was rejected in the present investigation in favor of a far more laborious experimental procedure which would permit accurate determination of these joint functional relationships.

[3] Williams (11) has applied the 3-dimensional design to an experimental situation and has pointed out the advantages of this design from the standpoint of methodology.

During extinction of the habit, the food release mechanism was disconnected, though the connection between the bar and the polygraph marker was retained so that the animal's reactions were still recorded. Drinking water was always available to the animal in the box.

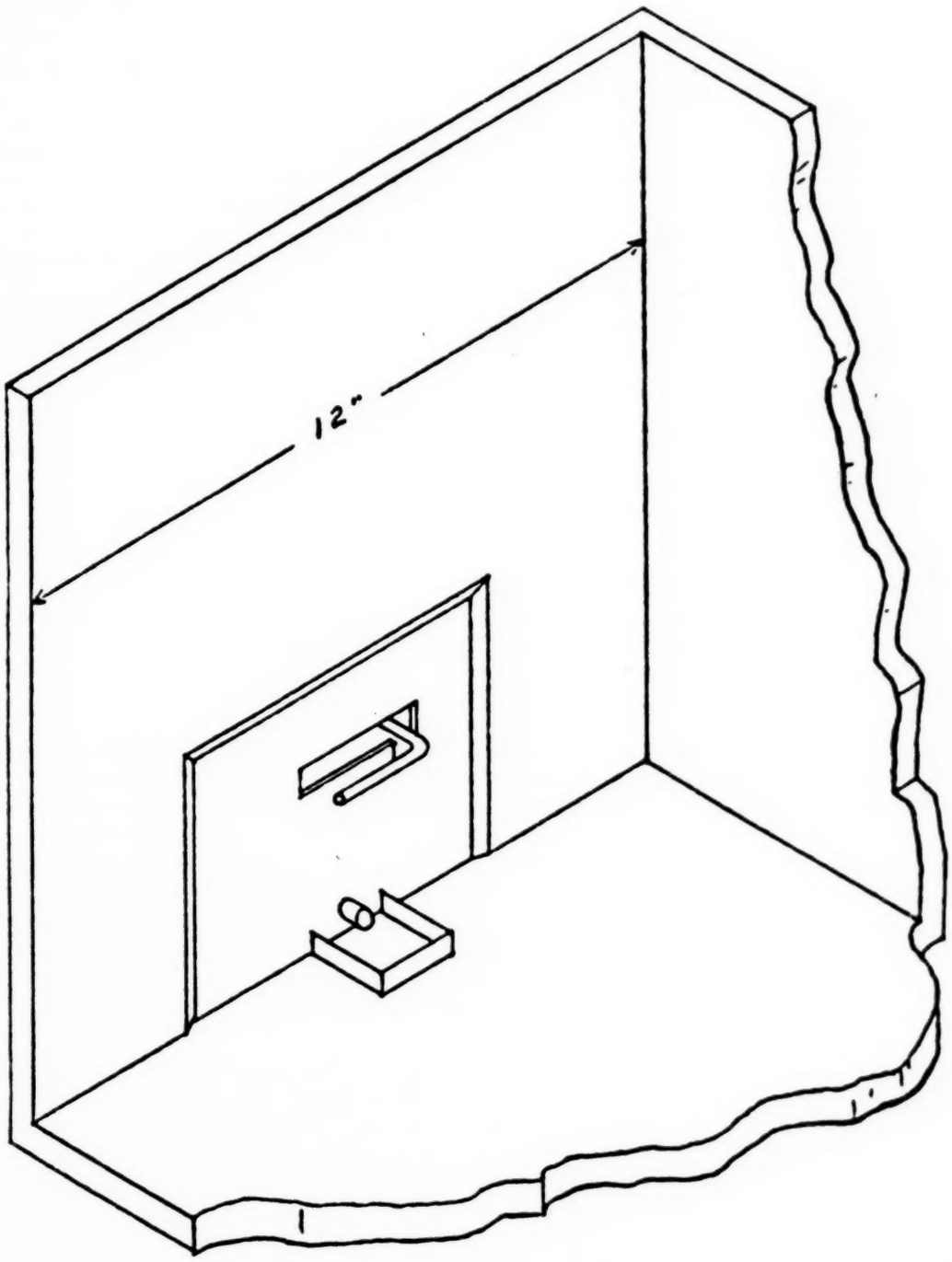

FIG. 1. Showing the interior of the experimental compartment with the horizontal bar inserted in position for pressing. Beneath the bar is the end of the glass tube through which pellets of food are delivered into the food tray.

PROCEDURE

Data were obtained from 40 rats in each of eight different subgroups employed. For convenience in experimentation and exposition, the animals were divided into two groups (A and B), and each of these main groups was in turn divided into four subgroups.

The two experimental variables selected for systematic investigation were: (1) the number of reinforced responses the animals were permitted to make during the learning period, and (2) the degree of hunger at the time the habit underwent extinction, the hunger being measured by the number of hours' time since the animal was satiated. These two antecedent variables are shown in Fig. 2. Actually the two main groups may be considered to make up two different experiments. Group A represents the animals which were used to determine the strength of the habit at the time of extinction as a function of the number of reinforced responses the animals were given during the learning. All of the animals in this group were extinguished with the same degree of hunger. Group B, on the other hand, represents animals which were all given the same number of reinforcements during learning but which were extinguished under different amounts of hunger. Thus Group B determines the strength of the bar-pressing response as a function of the degree of hunger at the time of extinction. By a fortunate circumstance, Williams (10) had already studied the extinction behavior of large groups of rats under conditions such that their results could be combined effectively with those of Groups A and B of the present experiment; we shall call the Williams animals Group W.

Figure 2 presents diagrammatically the subgroups of A, B, and W in relation to each of the two antecedent variables. A more precise numerical statement is given in Table 1. A study of

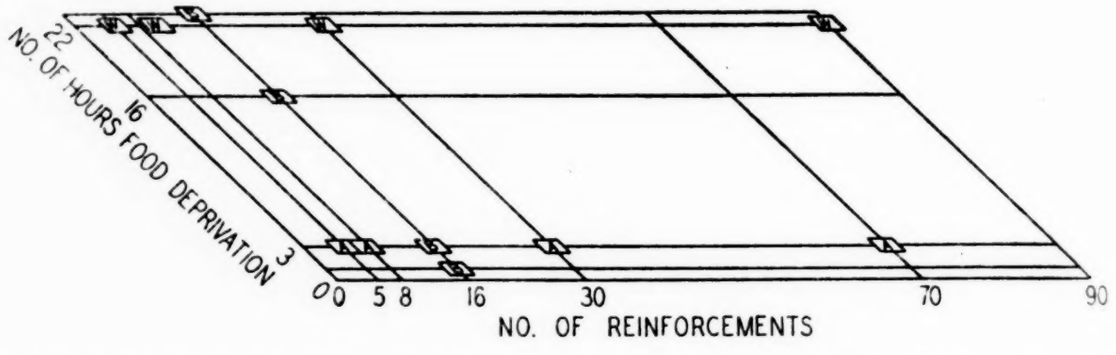

FIG. 2. Showing the different combinations of the two antecedent variables. The subgroups of Group A are represented by the letter A, those for Group B and the Williams experiment by the letters B and W respectively.

Figure 2 will show that Groups A and W are alike in that the number of repetitions was varied systematically, whereas extinction occurred under a constant number of hours of food deprivation in each (3 hours in Group A and 22 hours in Group W). In Group B, on the other hand, the number of hours of food deprivation was varied systematically while the number of reinforcements was kept constant at 16. Thus Group B, having one subgroup in common with Group A and one subgroup almost in common with Group W, constitutes a relational connecting link between the other two groups.

The general procedure involved in feeding the animals, habituating them to the experimental boxes, and training them to secure pellets by pressing the horizontal bar followed closely the precedent set by the experiments of Ellson (2) and Williams (10). Each animal was fed moist food for one-half hour daily for one week. The moist food was made by mixing a small amount of water with dry, finely ground calf meal to form a mash. In order to establish a definite feeding rhythm and to facilitate a more rapid satiation preceding extinction, this feeding occurred at the same time of day that the rat would later be trained and subsequently extinguished. Following this the animals were fed in the experimental compartments for two days, one-half hour each day, the food in this case being dry pellets which were readily accessible in the food tray of the box.

The experimental period proper consisted of three separate sub-divisions: (1) habituation to securing food in the food cup at the click of the food-release mechanism; (2) training on the bar-pressing habit; (3) extinction of the learned habit.

Habituation.—In the habituation to the click of the food mechanism, the animals were given 40 pellets one at a time, the food release being operated by the experimenter through the manipulation of a switch mounted on top of the experimental box. In this procedure, 3 pellets were present in the food tray when the rat was first placed in the box. As the third pellet was being eaten, another pellet was released into the food tray and the remaining pellets were given one at

TABLE 1

SUMMARY OF PROCEDURES FOR GROUPS A AND B AS WELL AS THE PROCEDURE FOR THE WILLIAMS EXPERIMENT

All learning was performed under 23 hours' food privation.

Subgroup	Number of Animals	Number of Reinforcements on Learning Day	Number of Hours' Hunger on Extinction Day
Group A			
1	40	5	3
2	40	8	3
3	40	30	3
4	40	70	3
Group B			
5	40	16	1
6	40	16	3
7	40	16	16
8	40	16	23
Williams' experiment			
1	35	5	22
2	35	10	22
3	35	30	22
4	35	90	22

a time just after the previous pellet had been completely consumed. This procedure served to prevent the rats from becoming frightened at the sound of the food mechanism, and the habit was set up whereby the animals obtained and ate a pellet immediately following each click of the mechanism. After the animals had been given 20 pellets in this manner they were returned to the living cages for 15 minutes, after which they were again put in the experimental boxes and given another 20 pellets in the same way.

Training.—The day following habituation to the boxes and to the food-delivery device, the animals were trained on the bar-pressing response. A small amount of moist food was placed on the brass panel immediately above the horizontal bar. In attempting to remove the moist food the animals pressed the bar accidentally and a food pellet was released into the tray. Usually three or four pellets were obtained in this 'accidental' fashion, and the first reinforced response was counted when the bar was pressed and the pellet eaten immediately after delivery.[4] The animals of the different subgroups were given the required number of pellet-reinforced responses in this manner, and were then returned to the living cages. When the data are examined for consistency within a subgroup and between subgroups for the time required to secure the first five pellets, the above procedure appears satisfactory (Table 2).[5]

A half hour after each rat had been returned to its living cage it was given the amount of dry food necessary to raise the total quantity eaten to 70 pellets for the session.

Extinction.—Since different subgroups of rats were to be extinguished with different degrees of hunger drive as measured by the number of hours of food deprivation, it was necessary to have all of the animals satiated to the same degree at the beginning of the privation period. Moist food was placed in the living cage one hour before the time from which drive was to be measured. Near the end of the one-hour satiation period each animal was fed by hand, both moist and dry food being presented. When an animal would no longer eat either the moist or the dry food even upon repeated presentations, the food was removed and the time noted in order to determine the hour at which the animal should be extinguished with the proper period of food deprivation.

[4] Animals engaged in removing moist food from the panel often did not find the pellets released by 'accidental' pressure on the bar until after a number of seconds had passed. The experimenter watched each animal closely in order to count as the first reinforcement that bar response which was followed immediately by the rat's eating of the pellet.

[5] Since one subgroup was allowed to make only 5 reinforced responses, the learning ability of all subgroups was compared on the basis of the first 5 reinforcements.

An arbitrary criterion of extinction was used, the bar responses being considered extinguished when a 5-minute period had elapsed during which no responses were made. Three measures of behavior potentiality were used. These were: (1) the number of extinction (non-rewarded) responses before the 5-minute criterion of no response; (2) the time required for extinction, exclusive of the 5-minute criterion; (3) the latency or time required to make the first 3 extinction responses, a single latency being the temporal interval between the point in time at which an animal was placed in the experimental compartment and that at which it pushed the bar.

TABLE 2

TABLE SHOWING EQUIVALENCE OF LEARNING ABILITY FOR THE VARIOUS EXPERIMENTAL SUBGROUPS

Group A		Subgroup Number			
		1	2	3	4
Time needed to secure first 5 pellets (minutes)	Mean	2.31	2.40	2.10	2.21
	Median	2.20	2.10	1.85	1.85
	$\sigma_{mdn.}$	.21	.21	.16	.21
Group B		5	6	7	8
Time needed to secure first 5 pellets (minutes)	Mean	2.37	2.15	2.24	2.32
	Median	2.20	1.95	2.10	2.00
	$\sigma_{mdn.}$	.28	.19	.19	.21

The time required to make the first response of the extinction series is the only latency measure not influenced by the extinction effect resulting from preceding non-rewarded trials, but in the present experiment three separate latencies were obtained from each animal at the beginning of the extinction period in order to insure more reliable latencies by means of larger sampling. The three latencies were separated by one-minute recovery intervals during which the animal was removed from the experimental box; these recovery intervals were introduced in the hope of eliminating, through spontaneous recovery, the major part of the extinction effects resulting from the preceding non-reinforcements. After the third latency response had been made, each animal was left in the experimental box until the criterion of 5 minutes with no response was reached.

During the extinction trials the food mechanism was disconnected and thus there was no characteristic click when the bar responses were made.

RESULTS

A preliminary picture of the nature of the present experimental results is yielded by the data distributions for resistance to experimental extinction as a function of reinforcements (food privation constant at 3 hours) as presented in Fig. 3. There it may be seen that the animals of the several subgroups, even with both motivation and number of reinforcements constant, varied considerably in their behavior potentiality. It is also apparent that all five of the distributions show a marked tendency to concentration at the zero extreme, this tendency being especially marked where the number of reinforcements is small. Finally, it is to be observed that the median values in all cases fall markedly below the means, exactly as is to be expected from the skewed nature of the distributions. This marked

skewness is characteristic of the other measures of behavior potentiality in Groups A, B, and W, and for this reason the median values have been selected as the basis for comparison of the various subgroups and in the later determinations of behavior potentiality in its functional relations with the antecedent variables, reinforcement and drive.

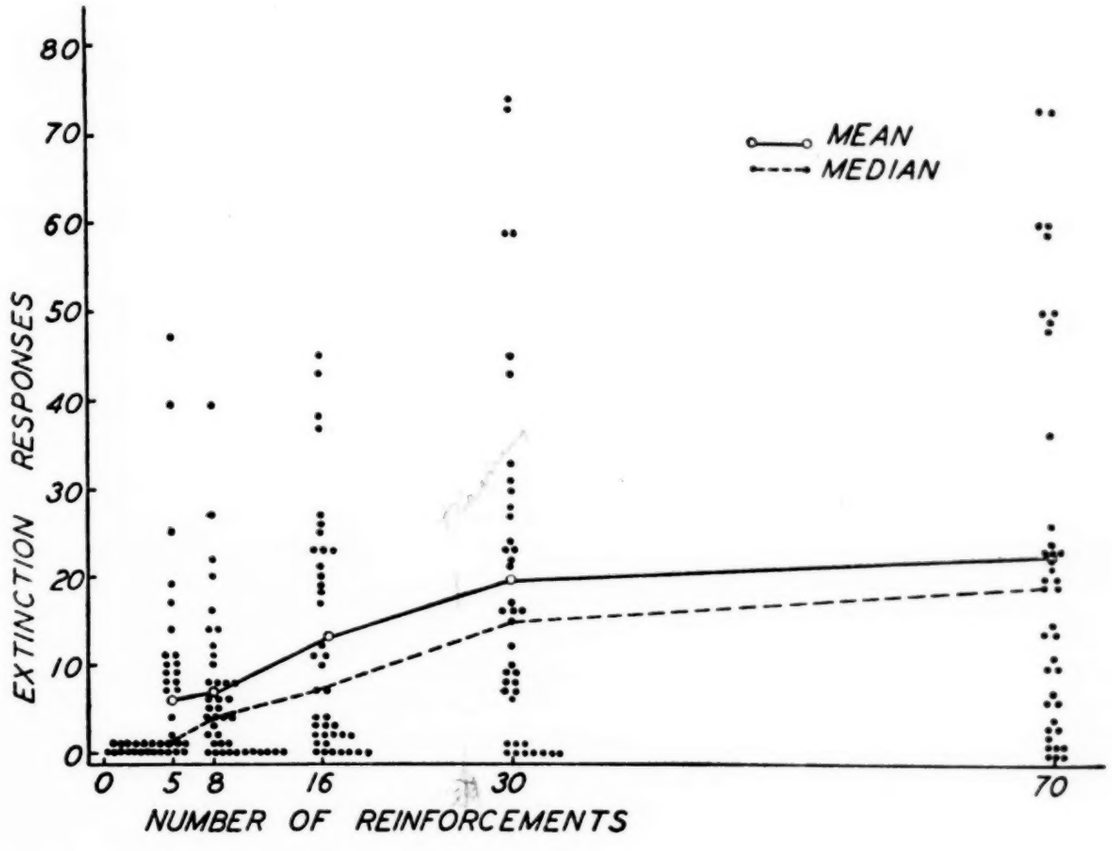

FIG. 3. Showing the distribution of the number of extinction responses of individual animals within the various subgroups for Group A. The solid line connects the means of the distributions and the dashed line shows the trend of the medians as reinforcement increases.

Behavior Potentiality—Group A

The numerical results of all three measures of behavior potential for all subgroups of Group A are given in Table 3, and the critical ratios indicating the reliabilities of the differences are given in Table 4. The results for the number of extinction responses obtained from Group A are shown graphically in Figs. 4, 6, 8, with the results for extinction time appearing in Fig. 9 and those for initial latency in Fig. 10. In general, an increase in the number of reinforcements with hunger drive constant for extinction, results in a corresponding increase in extinction responses and extinction time, with latency decreasing with increased reinforcement.

TABLE 3

SHOWING THE EFFECT OF THE NUMBER OF REINFORCEMENTS DURING TRAINING ON THE MEASURES OF BEHAVIOR POTENTIALITY

All animals had a 3-hour hunger drive at the beginning of the extinction period. $N = 40$ in each subgroup.

Group A

Measure of Behavior Potentiality		Number of Reinforcements			
		5	8	30	70
(*a*) Extinction responses	Mean	6.50	6.80	19.40	22.70
	Median	1.00	4.00	15.50	19.00
	$\sigma_{mdn.}$	2.07	1.66	3.91	4.88
(*b*) Extinction time (minutes)	Mean	4.95	4.45	8.13	8.94
	Median	0.75	3.25	7.05	9.45
	$\sigma_{mdn.}$	1.40	0.96	1.36	1.22
(*c*) Initial latency * during extinction period (minutes)	Median	2.33	0.90	0.60	0.50
	Q	2.20	1.31	0.66	0.95

* The initial latency was determined by taking three successive latencies for each of the 40 animals in a subgroup with a one-minute period between the latencies to permit spontaneous recovery. The 120 latencies for a subgroup were then combined into a single distribution and the median for that distribution was computed. Occasionally animals made no responses within the 5 minutes which were allowed and the latencies for those cases were indeterminate, making it impossible to compute means for the distributions. Also, it was considered unwise to compute the $\sigma_{mdn.}$ directly from the distribution in the way Garrett (6) outlines because of the scarcity of cases at the medians. In addition, no $\sigma_{m.}$ was available from which the $\sigma_{mdn.}$ could be computed. For these reasons the differences between subgroups were not tested for significance with regard to initial latency. Semi-interquartile ranges (Q) are given for initial latencies as an indication of variability.

TABLE 4

SHOWING THE RELIABILITIES OF THE DIFFERENCES BETWEEN THE MEASURES* OF HABIT STRENGTH. CRITICAL RATIOS (diff./$\sigma_{diff.}$) OF THE MEDIANS FOR PAIRED SUBGROUPS

Group A

Paired Subgroups (Number of Reinforcements)	Measures of Behavior Potentiality	
	Number of Responses	Extinction Time
5 reinf. vs. 8	C.R. = 1.13	C.R. = 1.47
5 reinf. vs. 16	1.82	2.11
5 reinf. vs. 30	3.28	3.23
5 reinf. vs. 70	3.40	4.68
8 reinf. vs. 16	0.98	0.85
8 reinf. vs. 30	2.71	1.66
8 reinf. vs. 70	2.91	4.00
16 reinf. vs. 30	1.82	1.51
16 reinf. vs. 70	2.56	3.31
30 reinf. vs. 70	0.56	1.31

* Critical ratios were not obtained for the initial latency measures because of the method used in securing the latencies (see footnote for Table 3).

Williams (**10**) reports that at 22 hours' hunger drive the number of extinction responses as a function of previous reinforcement conformed to the equation,

$$N = 62 - 62 \times 10^{-.017R}, \quad (1)$$

where N is the number of extinction responses, and R is the number of reinforced trials during acquisition. By the nature of equation (1), if R equals zero, N also equals zero, since that is the origin of the curve. The equation also indicates that the first reinforcement will yield a measurable amount of extinction response. Closer examination of Williams' data has since revealed that these were not valid assumptions, and consequently a better fit to the Williams data has been obtained with the equation,

$$N = 66(1 - 10^{-.018R}) - 4. \quad (2)$$

The curve obtained by substituting different R values in this equation is shown in Fig. 4. This is an exponential or growth curve which

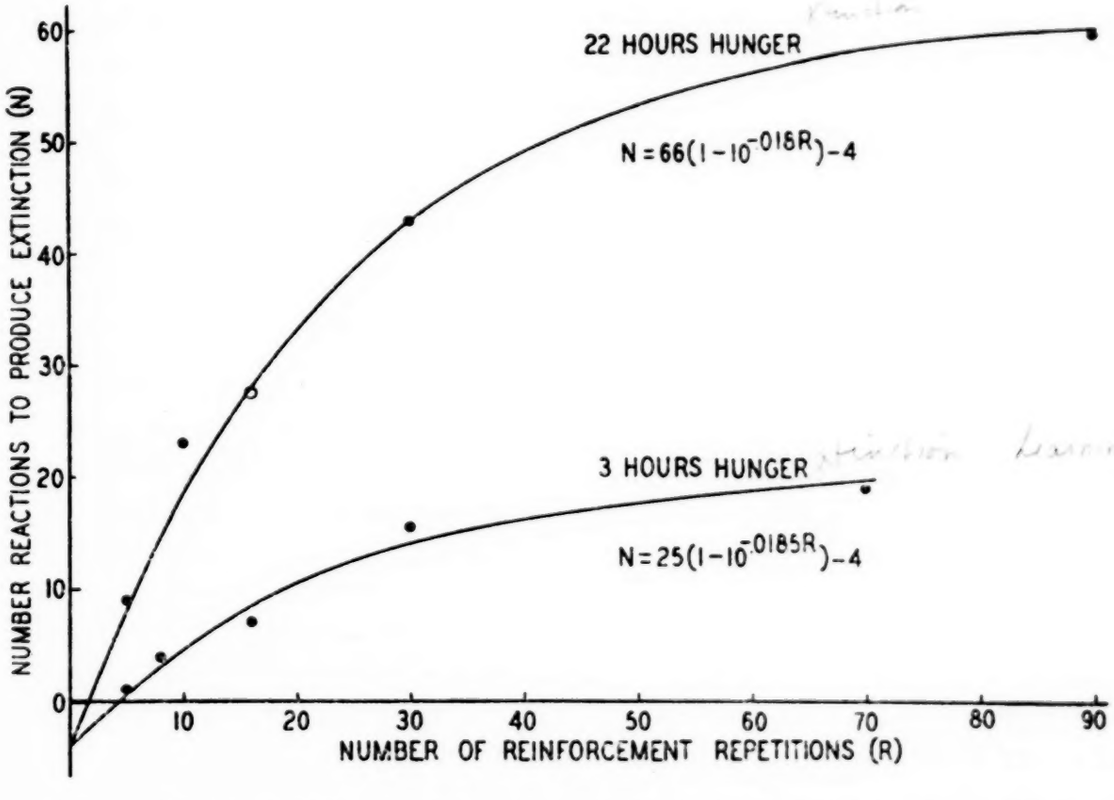

FIG. 4. Curves showing the relation of extinction responses (N) to the number of reinforcements (R) at two levels of drive. The upper curve of extinction at 22 hours' hunger is taken from Williams' experiment, while the lower curve at 3 hours' hunger represents Group A of the present experiment. The solid circles are the empirical values for the median number of extinction responses with the curves themselves being plotted by substituting different values for R in the corresponding equations. The hollow circle represents an interpolation in the equation of our Group B at $h = 22$ and $R = 16$. The close agreement of this value with Williams' results, as a whole, largely justifies the combination of the Williams' data with those of the present experiment.

approaches 62 extinction responses, the asymptote, when the number of reinforcements becomes very large. The curve also indicates that a minimum number of reinforced trials, in this case 1.51, must be given before the behavior potentiality becomes sufficiently large to result in an overt reaction and so in any number of extinction responses at all. The value 1.51 was determined by substituting $N = 0$ in equation (2) and solving for R. The -4 of equation (2) indicates that absolute zero lies 4 extinction units below the reaction threshold.

The same general type of equation (exponential) was fitted to the data for Group A of the present experiment, and is represented graphically as the lower curve in Fig. 4. This equation is

$$N = 25(1 - 10^{-.0185R}) - 4. \quad (3)$$

From examination of the two curves shown in Fig. 4, it is evident that the degree of hunger plays an important role in determining the characteristics of those curves. Since the degree of hunger at the time of extinction is the only significant antecedent variable by which the two curves can be differentiated, it is evident that a change in drive changes the ultimate limit of the curve, but the origin does not change and the coefficients of R in the two equations (.018 vs. .0185) are practically identical.

Comparable curves were fitted to the data for the time necessary to produce extinction and to the initial latency values during the extinction period (Figs. 9 and 10).

Behavior Potentiality—Group B

It is evident from a comparison of Williams' results with those from Group A (Fig. 4) that there is a marked positive relation between the drive at the time of extinction and the resistance to extinction. The animals of Group B were run primarily in order to secure quantitative statements of the nature and extent of this relationship.

The results of the three measures of behavior potentiality for Group B are given in Table 5, and the critical ratios indicating the reliabilities of the differences appear in Table 6. The results for the number of extinction responses are shown graphically in Figs. 5 and 8, with those for extinction time and initial latency appearing in Figs. 9 and 10 respectively. Equations have been fitted to the data for Group B, as in Group A. For example, when the number of reinforcements is held constant, the equation for the number of extinction responses as a function of the number of hours' hunger (h) at the time of extinction becomes

$$N = 9.4 \times 10^{.0241h} - 4. \quad (4)$$

TABLE 5

SHOWING THE EFFECT OF THE DEGREE OF HUNGER AT THE TIME OF EXTINCTION ON THE MEASURES OF BEHAVIOR POTENTIALITY

All animals had been given the same amount of training (16 reinforcements). $N = 40$ in all subgroups.

Group B

Measure of Behavior Potentiality		Number of Hours' Hunger			
		1	3	16	23
(*a*) Extinction responses	Mean	9.10	12.30	24.50	32.90
	Median	6.00	7.00	19.00	29.50
	$\sigma_{mdn.}$	2.01	2.57	4.34	4.46
(*b*) Extinction time (minutes)	Mean	4.61	6.24	9.76	12.84
	Median	4.45	4.45	7.65	13.75
	$\sigma_{mdn.}$	0.44	1.05	1.53	1.09
(*c*) Initial latency * during extinction period (minutes)	Median	0.95	0.90	0.45	0.25
	Q	1.74	1.52	0.42	0.18

* See footnote for Table 4.

The curve shown in Fig. 5 indicates that even though the animals have a zero hunger drive, as measured by food deprivation, they may be expected to make a number of non-reinforced responses. Corroboration of this conclusion has been given by Skinner (**9**), who has demonstrated that animals will make bar-pressing responses when there is zero food deprivation (complete satiation), and by Finch (**3**), who has demonstrated that conditioned salivary responses occur in satiated dogs.

The curve in Fig. 5 would lead one to expect that as hunger increases indefinitely the behavior potentiality will also increase. Ob-

TABLE 6

SHOWING THE RELIABILITIES OF THE DIFFERENCES BETWEEN THE MEASURES* OF BEHAVIOR POTENTIALITY. CRITICAL RATIOS (diff./$\sigma_{diff.}$) OF THE MEDIANS FOR PAIRED SUBGROUPS

Group B

Paired Subgroups (Hours of Hunger)	Measures of Behavior Potentiality	
	Number of Responses	Extinction Time
1 hour vs. 3	C.R. = 0.31	C.R. = 0.00
1 hour vs. 16	2.72	2.01
1 hour vs. 23	4.81	7.88
3 hours vs. 16	2.38	1.72
3 hours vs. 23	4.37	6.16
16 hours vs. 23	1.69	3.25

* See footnote for Table 4.

viously this cannot be the case, since eventually the effects of starvation will begin to act adversely on the behavior potentiality and as the animal nears death from starvation it will be able to make progressively fewer bar-pressing responses. Skinner (9) has obtained a curve of habit potentiality extending out to about six days of food deprivation. His graph shows that behavior potentiality rises with a slight negative acceleration to a maximum at approximately five days, and thereafter declines sharply, many of the animals dying by the sixth day of starvation. Finch (3) has suggested that the curve

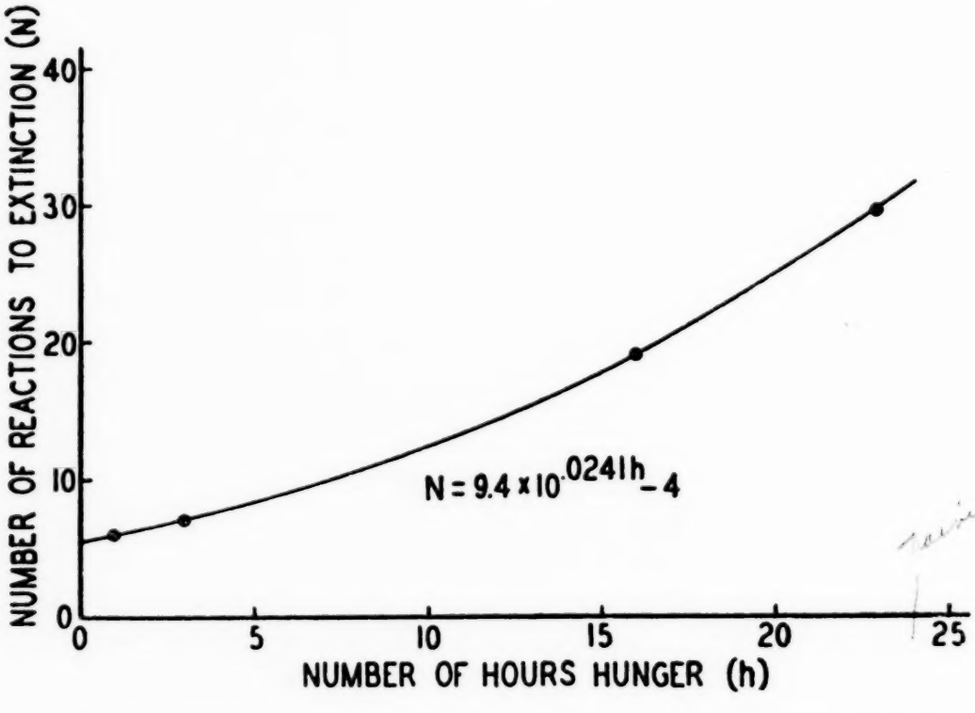

FIG. 5. Curve showing the number of extinction responses (N) as a function of hunger (h) when the number of previous reinforcements is constant at $R = 16$. The circles represent the empirical median values from Group B. The curve was plotted by substituting different values for h in the equation shown beneath the graph.

for the conditioned salivary response in dogs follows a simple parabolic function, rising with a slight negative acceleration to a maximum at about 72 hours of food deprivation and thereafter declining. The above considerations obviously place marked limitations on the generality of equation (4) with its indications of a positively accelerated rise from zero to 23 hours' hunger. Unfortunately, no points between zero and 24 hours are available either in Finch's study or in the related one of Skinner (9). However, the fact that the number of responses at 24 hours' deprivation represents a distinct 'hump' in Skinner's curve between zero and 48 hours' deprivation suggests the possibility of a positively accelerated curve existing between zero and 24 hours in that study.

But why should there be a positive acceleration of behavior potentiality in this particular region of food deprivation which, upon the whole, seems to be negatively accelerated during the first few days? The following tentative hypothesis is offered: in both the Skinner study and the present one, the original reinforcement was carried out at close to 24 hours of hunger. Presumably the hunger stimulations characteristic of this amount of food deprivation make up an important portion of the stimulus complex conditioned to the bar-pressing reaction. Now it has been shown by Hovland (**8**) that *intensity* of tonal stimulation is generalized.[6] Moreover, it has been shown (**7**) that *qualitative* generalization follows a marked negatively accelerated falling gradient from the point on the stimulus continuum at which conditioning occurred. It therefore is plausible to assume that there would be a negatively accelerated falling gradient of behavior potentiality from the degree of hunger actually conditioned to any other degree of hunger at which extinction may occur, if no other behavior tendencies were present. On these assumptions it is reasonable to suppose that there would be a typical generalization gradient of behavior potential extending in each direction from 24 hours' food deprivation. This double-winged gradient superposed upon a mildly negatively accelerated curve suggested by the early portions of Skinner's graph (**9**) would produce the 'hump' in Skinner's curve at 24 hours, and the positive acceleration of Fig. 5 which extends from 0 to 23 hours. This is to say, in effect, that while equation (4) probably represents accurately the experimental results from 0 to about 24 hours under the conditions of both Skinner's experiment and our own, both results are presumably somewhat distorted by generalization effects.

Behavior Potentiality as a Joint Function of Habit Strength and Motivation

In the statement of results so far, the two groups (A and B), together with Williams' data, have been considered separately. Of far greater importance from the point of view of the present investigation is the manner in which the design of the experiment permits these discrete bits of data to be colligated and treated as a multivariable investigation.

Figure 6 is an orthographic projection with the two independent variables of the investigation occupying the two dimensions of the base and a measure of behavior potentiality (the dependent variable)

[6] Hovland (**8**) has presented evidence for the generalization of intensity of the conditioned stimulus in an experiment using the conditioned galvanic response. He found that the strength of the conditioned response varies inversely with the remoteness of the stimulus from the point which was originally conditioned.

represented by columns rising perpendicularly in the third dimension. The shaded columns represent the empirical values for the Williams experiment. Moreover, in Fig. 6 the justification for combining the Williams experiment with the results of the present investigation begins to appear. Not only does the curve at 22 hours' hunger parallel the exponential curve for extinction responses at 3 hours' hunger without any significant inversions, but also the results of Group B serve as an important indication of the consistency of this

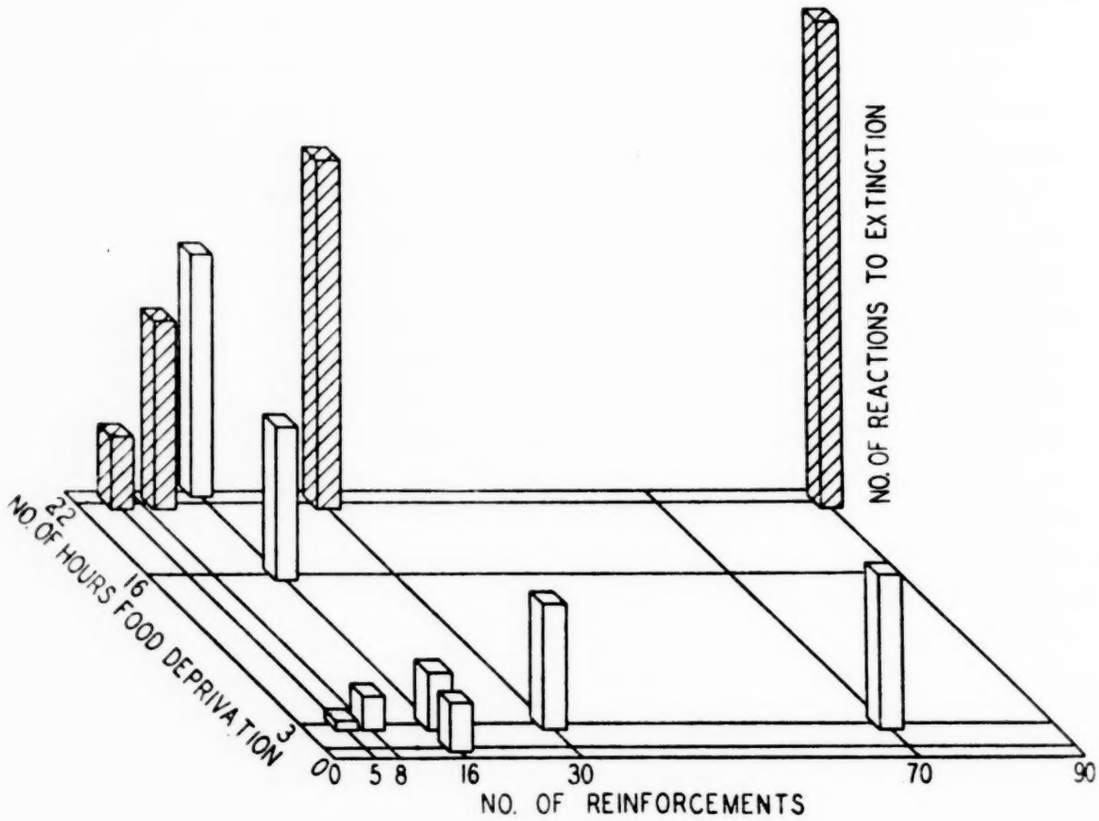

FIG. 6. Three-dimensional representation of the empirical values, the median number of reactions to extinction. The shaded columns are Williams' results; the columns from left to right along the line for 3 hours' food deprivation represent Group A; and the columns running diagonally along the line of 16 reinforcements in the direction of increased deprivation are Group B.

relationship. When the value $h = 22$ is substituted in the equation fitted to Group B (equation 4) the resulting N value falls almost exactly in line with the curve previously fitted to Williams' results. This interpolated N value from equation (4) is represented by a hollow circle in the graph of the Williams data (Fig. 4). Because of this striking consistency in the number of extinction responses yielded by the two experiments, it has been possible to utilize the Williams data in the mathematical analysis with a confidence otherwise unwarranted.

As pointed out previously, the results of Group A and those of the Williams experiment indicate that behavior potentiality is a simple

growth function of the number of reinforcements when hunger at the time of extinction is constant. This coupled with the fact that Group B shows the relation of behavior potentiality to food deprivation, makes it possible to fit a geometrical surface that takes into account all three variables—reinforcement, hunger, and the measure of behavior potentiality.

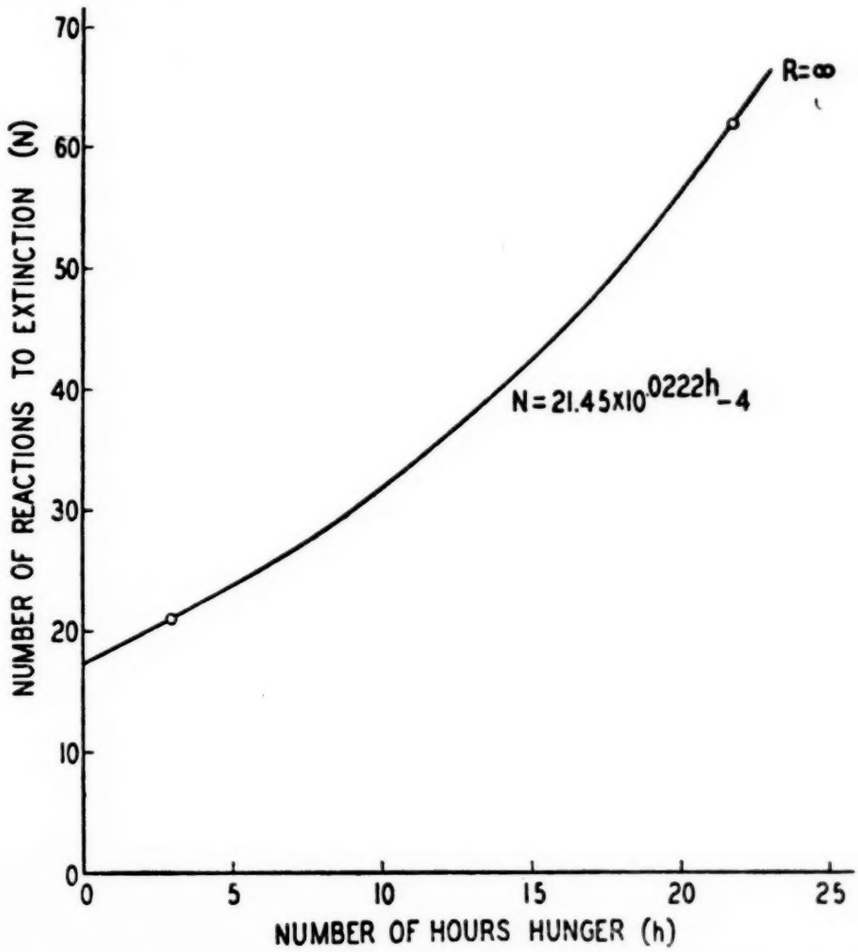

FIG. 7. Hypothetical curve of the number of extinction responses as a function of hunger when the number of reinforced training trials is infinite. The lower circle represents the probable maximum number of extinction responses for Group A with an infinite number of reinforcements, and the upper circle represents the probable number of extinction responses for Williams' data under the same condition of infinite reinforcement.

An equation representing the results presented in Fig. 6 as a curved surface has been derived in the following manner. First, the two growth curves along the reinforcement dimension were extrapolated by substituting $R = \infty$ in equations (2) and (3), in order to find the ultimate limits of the curves which might be expected to obtain from a very large number of reinforcements. Then by assuming that these and all other comparable asymptotes between $h = 0$ and $h = 23$ conform to the same general function of the amount of hunger drive as that fitted to Group B at 16 reinforcements (equa-

tion 4), the equation for the asymptotes for the exponential curves was found to be,

$$N = 21.45 \times 10^{.0222h} - 4. \quad (5)$$

This presumptive curve for the asymptotes is shown in Fig. 7. Finally, by substituting this expression of the generalized asymptotes as a function of drive for the asymptotes in equations (2) and

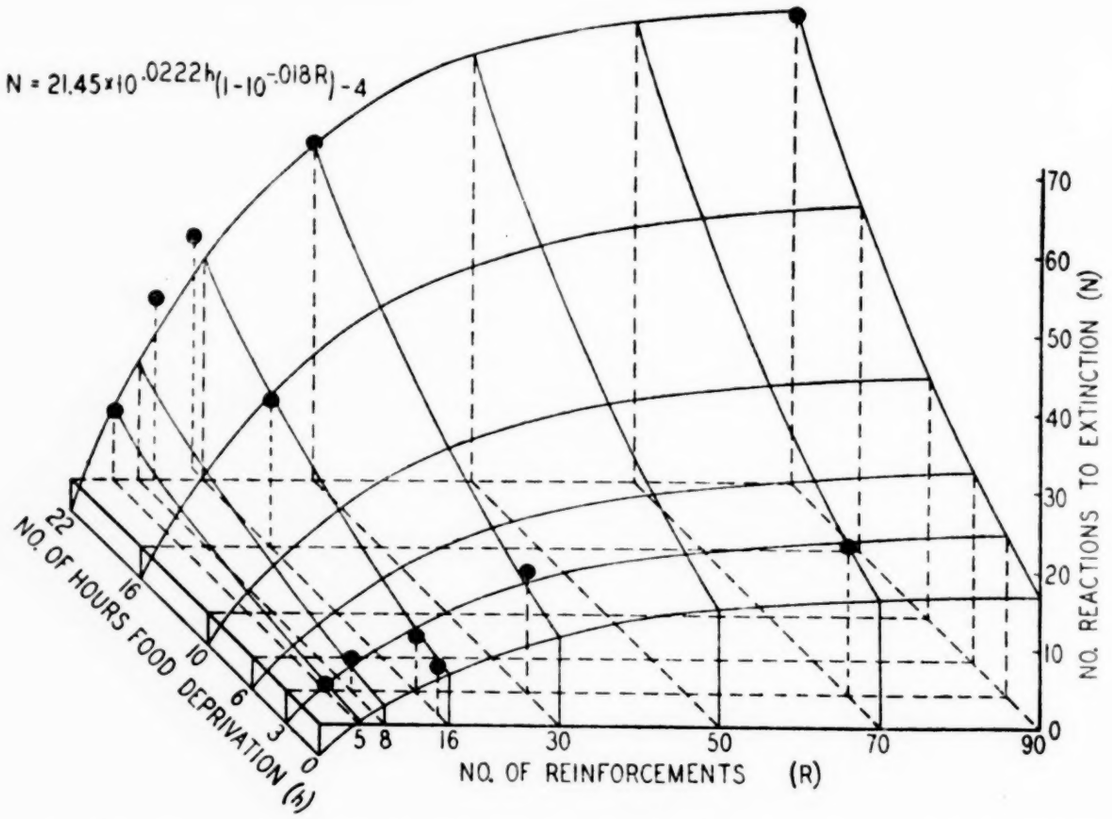

FIG. 8. Surface fitted to the data in Fig. 6 and Tables 3 and 5. The solid circles represent the empirical values and correspond to the tops of the columns in Fig. 6. The solid curved lines are derived by substituting in the general equation for the surface and are inserted to aid the eye in following the curvature of the surface with changes in food deprivation and reinforcement. The vertical dashed lines are projections from the surface to the plain. The fitted equation describing this surface is in the upper left-hand portion of the figure.

(3), which are otherwise practically identical, the mathematical expression for the surface was found to be

$$N = 21.45 \times 10^{.0222h}(1 - 10^{-.018R}) - 4. \quad (6)$$

By substituting a number of different combinations of h and R in the above equation and solving for N, enough values of N were easily obtained to permit accurate plotting of the smooth surface, as shown in Fig. 8.

Inspection of Fig. 8 indicates that the conclusions which have already been drawn for individual groups (A and B) should hold

approximately for intermediate combinations of hunger and reinforcement if those combinations were subjected to test. Here may be observed the characteristic exponential course that behavior potential values take when the number of reinforcements is increased with drive constant. Here also may be observed the course that behavior potential values take for any given number of hours' hunger (between zero and 22) at the time the response is extinguished. Figure 8 also shows plainly the potential remaining when there is zero hunger. This residual potential turns out to be a definitely

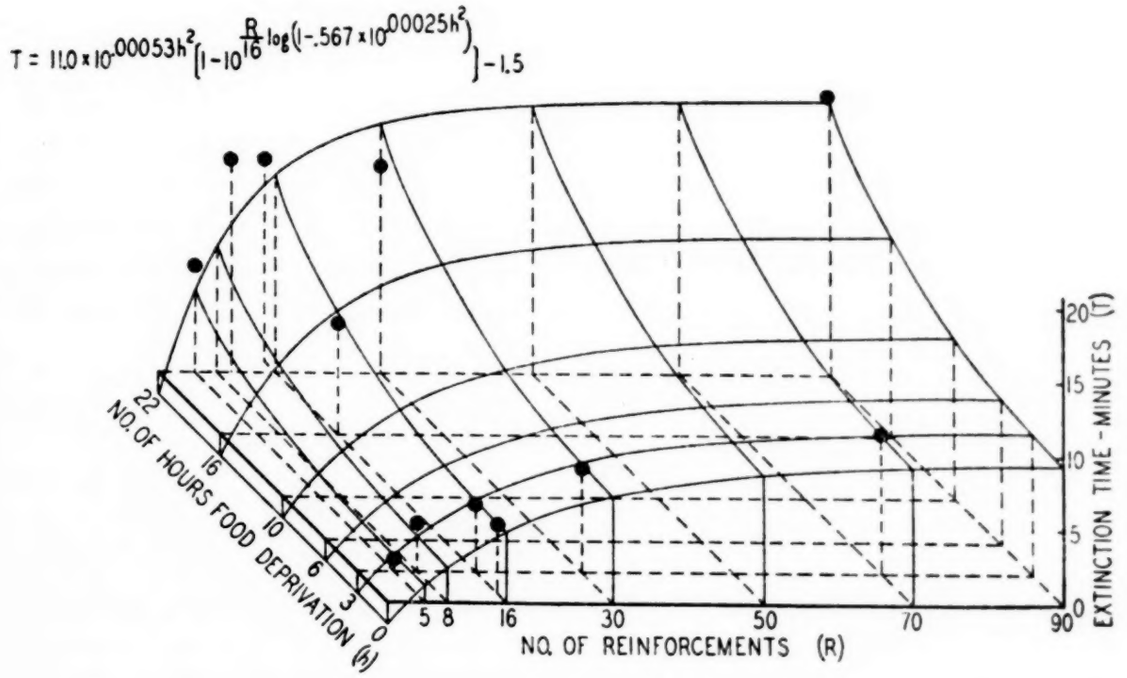

FIG. 9. The surface for time to extinction as a joint function of hunger and the number of training reinforcements. The equation describing the surface is shown in the drawing.

positive growth function of the number of reinforcements, with small but finite values. When the number of reinforcements becomes zero, however, the number of responses falls to -4, which represents the distance of absolute zero below the reaction threshold.

Since the surface in Fig. 8 purports to be a graphic presentation of behavior potentiality as a function of both hunger and reinforcement, the equation for this surface (equation 6) enables us to make a verbal statement of this relationship. Hunger (h) and number of reinforcements (R) appear in exponents within equation (6), with the separate quantities in which h and R appear subsequently being combined as a product. Therefore, the statement can be made that *functions* of h and R combine multiplicatively in their determination of the number of extinction responses as a measure of behavior potentiality. The indication is, further, that the function of h (what-

ever it ultimately turns out to be when cleansed from the presumptive contamination of generalization effects) is the asymptote of the expression of which the number of reinforcements is the exponent.

The surface for the time necessary to produce extinction (Fig. 9) has the same general characteristics as the surface for the number of extinction responses, although a more complicated equation was required to yield a reasonably good fit. The equation for the time necessary for extinction as a joint function of previous reinforcement and hunger when the habit underwent extinction is

$$T = 11.0 \times 10^{.00053h^2}\left[1 - 10^{\frac{R}{16}\log(1-.567\times 10^{.00025h^2})}\right] - 1.5. \quad (7)$$

The important difference between this equation and the one for number of extinction responses (equation 6) is the fact that food deprivation in equation (7) determines in part the rate of rise of the exponential curves with increased reinforcement. It will be recalled that in the surface for the number of extinction responses, the factor of drive affected only the ultimate limits or asymptotes of the ex-

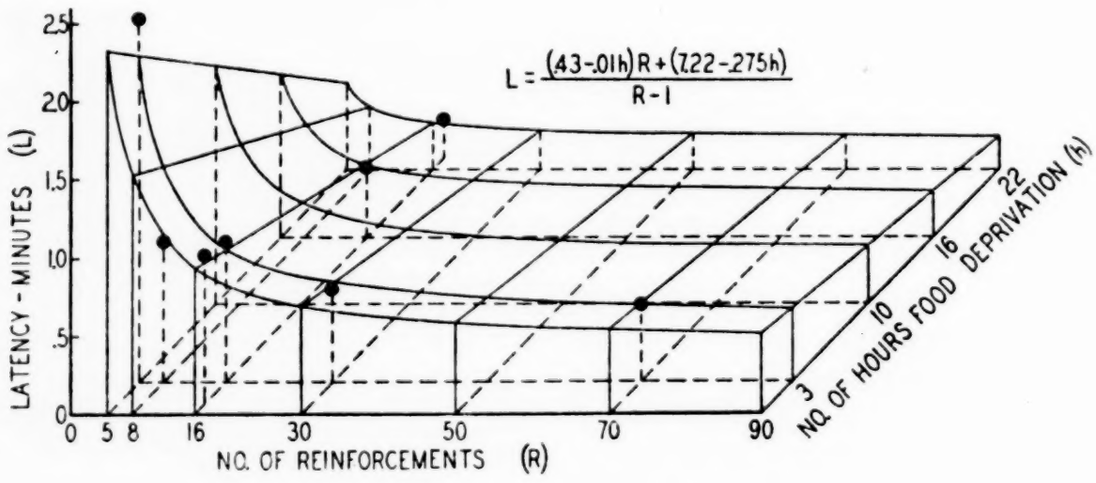

FIG. 10. Surface showing latency as a joint function of hunger and number of training reinforcements. The equation for the surface is shown with the drawing. Comparison with Figs. 8 and 9 shows the inverse relation of latency to the other two measures of behavior potentiality.

ponential curves, the exponent determining the rates at which these curves approached their asymptotes remaining the same regardless of the drive with which the exponential curve was secured. Certain irregularities in the Williams data on time to produce extinction suggest that equation (6) and Fig. 8 may present a truer picture of the functional relationship here under investigation than do equation (7) and Fig. 9. This question can be decided only by further experimentation.

Despite the fact that the Williams investigation yielded no data for latency, a surface showing latency as a function of drive and reinforcement has been constructed analogous to Figs. 8 and 9 on the tentative assumption that the general type of curve found for the series of latencies at 3 hours' hunger will also hold for latency at zero and 23 hours' hunger as reinforcements increase. This assumption was arrived at by analogy with the surface for the number of extinction responses in which the exponent determining the rate of change for the reinforcement curve remained unchanged regardless of the amount of food deprivation. When latency is used as a measure of behavior potentiality, a rather different picture results (Fig. 10). This difference arises from the well known fact that latency has an inverse relationship to the other two measures; *i.e.*, the stronger the action potentiality, the shorter the time necessary for a given response to take place when the opportunity presents itself. This relationship may be seen in the drawing of the fitted latency surface (Fig. 10) and in the equation for latency as a function of food deprivation and reinforcement. This is

$$L = \frac{(.43 - .01h)R + (7.22 - .275h)}{R - 1}. \qquad (8)$$

Among other things, the surface and its equation indicate that whenever the number of reinforcements is one or less, the initial latency during extinction will be infinitely great. This is, of course, not surprising, since the surface for extinction responses indicates that no responses will occur when there has been no previous reinforcement (or only one reinforcement); it therefore follows logically that the latency for no response will be infinite. Also, it is to be noted that a substantially linear relation exists between latency and hunger drive as opposed to the positively accelerated curves which were found for extinction time and number of responses as functions of hunger. It must again be emphasized that extrapolation is precarious, especially in the direction of increased hunger, because it would appear from equation (8) that when there are 34.2 hours of hunger with 16 reinforcements having been given previously, the time necessary for an animal to press the bar is zero minutes (latency = 0). Obviously a zero latency is a physical impossibility.

Summary

1. The experiment was designed to investigate the joint effect of two antecedent variables on three measures of behavior potentiality. The two antecedent variables were: (1) the number of reinforced repetitions during acquisition of a bar-pressing habit, and (2) the

degree of hunger drive as measured by length of food deprivation, when the bar-pressing response underwent extinction. The measures of behavior potentiality were (*a*) the number of unreinforced responses necessary to produce extinction, (*b*) the length of time required to extinguish the response, and (*c*) the latency of the response at the beginning of the extinction period.

2. The subjects were 320 albino rats divided into 8 subgroups of 40 rats each. Four of these subgroups were given 5, 8, 30, and 70 reinforced training trials respectively, and the behavior potentiality was measured when all the animals had a 3-hour hunger drive. The remaining four subgroups were all given 16 reinforced trials, but had 1, 3, 16, or 23 hours' hunger drive at the time the bar-pressing response was extinguished.

3. The results of an earlier experiment by Williams (**10**) were found to be quantitatively comparable with the results for the number of extinction responses in the present experiment, and are included in the data from which conclusions are drawn. Williams' results on the time necessary for extinction, while manifesting certain internal inconsistencies as well as deviations from the results of the present experiment, are incorporated in the present investigation though with some reservations.

4. The joint determination of the behavior potentialities by the two antecedent or independent variables can be adequately represented only by means of equations or by curved surfaces. Such equations, accompanied by the curved surfaces corresponding to them, are presented for each of the three measures of behavior potentiality. Of the three, the data for the number of reactions to produce extinction are believed to be the most adequate.

5. A study of these three multiple relationships discloses a striking similarity between the equations and surfaces representing the number of reactions to produce extinction and the time required to produce extinction as joint functions of the antecedent variables (reinforcements and food deprivation). Further experimentation will be necessary to determine whether the differences between equations (6) and (7) are genuine or whether they are due to limitations in the sample. The combination of reinforcements and food deprivation (*habit strength* and *motivation*) to yield values of reaction latency results in a surface and equation of radically different form than those for extinction responses and time.

6. If the equation for the number of reactions required to produce extinction be taken as the true relationship (assuming that the true equation for time to produce extinction would differ only in the value of one or more of the constants) a fairly simple verbal statement of the relationship may be made: behavior potentiality increases as a

positive growth (exponential) function of the number of reinforcements, the asymptote of the growth function being itself a function of the number of hours' food deprivation. The exact functional relationship of this asymptote to motivation in general or even to the number of hours' food deprivation is not clear, though the probability is that it is a slightly negatively accelerated increasing function for deprivation less than 5 or 6 days.

7. However, if the exact mathematical statement be waived, a statement of the relationship may be made with less definiteness but with far greater confidence: (*A*) The number of extinction reactions and the time required for extinction are both *increasing* functions of the number of previous reinforcements and the number of hours' food deprivation (from zero to about 120 hours). (*B*) Reaction latency is a *decreasing* function of the same two antecedent variables.

(Manuscript received June 18, 1941)

References

1. Crutchfield, R. S., Efficient factorial design and analysis of variance illustrated in psychological experimentation, *J. Psychol.*, 1938, **5**, 339–346.
2. Ellson, D. G., Quantitative studies of the interaction of simple habits. I. Recovery from specific and generalized effects of extinction, *J. exp. Psychol.*, 1938, **23**, 339–358.
3. Finch, G., Hunger as a determinant of conditional and unconditional salivary response magnitude, *Amer. J. Physiol.*, 1938, **123**, 379–382.
4. Fisher, R. A., *Design of experiments*, Edinburgh: Oliver and Boyd, 1937.
5. Fitts, P. M., Perseveration of non-rewarded behavior in relation to food-deprivation and work-requirement, *J. genet. Psychol.*, 1940, **57**, 165–191.
6. Garrett, H. E., *Statistics in psychology and education*, New York: Longmans, Green and Co., 1937.
7. Hovland, C. I., The generalization of conditioned responses. I. The sensory generalization of conditioned responses with varying frequencies of tone, *J. gen. Psychol.*, 1937, **17**, 125–148.
8. ——, The generalization of conditioned responses. II. The sensory generalization of conditioned responses with varying intensities of tone, *J. genet. Psychol.*, 1937, **51**, 279–291.
9. Skinner, B. F., *The behavior of organisms*, New York: D. Appleton-Century Co., 1938.
10. Williams, S. B., Resistance to extinction as a function of the number of reinforcements, *J. exp. Psychol.*, 1938, **23**, 506–522.
11. ——, Transfer of extinction effects in the rat as a function of habit strength, *J. comp. Psychol.*, 1941, **31**, 263–280.

CONDITIONING OF EYELID CLOSURE WITH VARIOUS CONDITIONS OF REINFORCEMENT [1]

BY HENRIETTA WEBER

Philadelphia

AND

G. R. WENDT

Wesleyan University, Middletown, Conn.

Introduction

This study is concerned with a comparison of conditioning of eyelid closure in humans under four conditions of reinforcement: (1) a light stimulus followed by a loud sound, (2) light followed by a puff of air to the eye, (3) light followed either by a loud sound or an air puff in an unpredictable order, (4) light followed by simultaneous loud sound and air puff. A separate group of subjects was used for each condition of reinforcement. The results of the investigation indicate that the loud sound is the least effective for the development of conditioned eyelid closures; simultaneous stimulation by sound and air puff is most effective; while puff alone and random alternation of sound or puff are intermediate in effectiveness and approximately equivalent to one another.

One may be interested in at least four aspects or interpretations of the data: (1) a comparison of the rate of conditioning when using a loud sound and when using an air puff as the conditioned stimuli, and speculation concerning the reasons for the greater effectiveness of the puff; (2) comparison of the effects of simultaneous sound and puff stimuli with the effects of either used singly (the greater effectiveness of the former is presumably a function of greater intensity); (3) comparison of the group where heterogeneous stimulation (sound and puff alternating) was used with that where homogeneous stimulation (sound alone or puff alone) was used. (Various studies of the work decrement have shown that the development of decrement is sooner, and the total work done is less, when the task is homogeneous (**11**). Conditioning procedures with massed homogeneous reinforcement may, if conditioning and work experiments are similar, favor the development of decrement. Hovland (**8**) has called such decre-

[1] This investigation was done in the Psychological Laboratory of the University of Pennsylvania in 1938–39.

ment during reinforcement the 'inhibition of reinforcement.' It is said to be revealed by the fact that a rest interval will be followed by increased responsiveness.); (4) attention to changes in the amplitude of the reflex eyelid closures to the light stimulus as a result of the conditioning procedures (there has been some disagreement among investigators as to whether reflex 'sensitization' occurs).

Apparatus and Procedure

The general features of our method were similar to those of the investigations of Hilgard and Marquis and others (7) on conditioning of eyelid response. The seated subject, with head steadied in the upright position, was arranged for photokymographic recording of movements of the upper eyelid. With each ready-signal she (all subjects were women) looked at a fixation light. A sudden increase in the brightness of this light was the conditioned stimulus. It was followed after 0.4 second by a conditioning stimulus (loud sound, puff of air, or both). Stimuli were repeated at 20 second intervals. Certain details of our apparatus and procedure differed from previous investigations.

(*a*) *Posture of Subject.*—The subject was seated in a straight chair with her head cushioned at the rear against a felt-padded brace, cut to conform to the head, and adjustable for height. The head was further steadied by a biting-board. The subject's arms rested on a table which was placed immediately in front of her. Her vision was unobstructed.

(*b*) *Recording and Stimulus Control.*—We employed a slit-type, bifilar-pendulum-photokymograph (13). It used $2\frac{1}{2}'' \times 7''$ sensitive paper; its speed such that the 7-inch record represented 830 ms. The pendulum in its swing tripped contacts which operated remote relays for presentation of the stimuli. The visual stimulus appeared about 25 ms after the beginning of the record, the conditioning stimuli 400 ms later.

The movements of the upper lid were recorded by the mirror method. A light bamboo lever, 5 inches long, 0.38 grams in weight, was attached at one end to the upper lid, directly above the lashes, by means of adhesive, at the other end to a fulcrum. A concave mirror, mounted at the fulcrum, formed an image of a straight lamp filament across the vertical slit of the photokymograph. The mirror was 10 inches from the slit; vertical movements of the lid therefore produced corresponding vertical movements of the filament image at the slit, magnified 4 times. A new method of mounting the recorder (not now described) allowed considerable lateral, forward, or rotational head movement without discomfort to the subject or disturbance of the record.

(*c*) *Visual Stimulation.*—Five feet directly in front of the subject was a 30-inch square, white-painted background, illuminated by a lamp shielded from the subject. Its surface brightness was about 2 apparent foot-candles. This was intended to prevent large sensitivity changes due to dark adaptation. The remainder of the room was dark. In the center of the background was a $\frac{3}{4}$ inch circular aperture in which was translucent white typewriting paper. Directly behind this was a flashlight lamp (6.2 v., 0.31 amp.), kept at a low level of incandescence as a fixation light, increased to full brilliance as a stimulus light.[2] This increased brilliance, the conditioned stimulus, lasted in each case for 650 ms.

(*d*) *Auditory Stimulation.*—The loud sound was produced in a headset consisting of two magnetic diaphragm telephone receivers. Ten volts of 1000 cycle alternating current was impressed for 150 ms. The resulting sound was unpleasantly loud. From experience with similar phones, one may estimate the sound to have been about 95 db above the threshold of audibility.[2]

(*e*) *Air Puff.*—The air puff was administered in the region of the external canthus of the left eye. The fall of a column of mercury in a manometer, magnetically released, furnished the power. The stimulus intensity was sufficient to elicit almost complete eyelid closures without secondary or maintained closure.

(*f*) *Procedure with Each Subject.*—Each subject served for one period during which she was given 50 double stimulations separated by intervals of about 20 secs. There was a longer pause between trials 25 and 26 while the photokymograph paper-holder was reloaded. (There was

[2] Due to a misunderstanding, the apparatus was dismantled before photometric and audiometric determinations were completed.

neither an adaptation series nor an extinction series, which have been customary in previous investigations.) The subject was told that there would be loud sounds, or puffs of air, or both, but not informed of their relationship to the light stimulus. She was told to sit quietly and to look at the fixation light when the ready-signal was given. The stimulation followed the ready-signal by about 3 secs. Other routine operations produced noises, such as those incidental to the removal of exposed papers, which served as cues supplementary to and preceding the ready-signal. Subjects were told to relax between trials.

(*g*) *Plan of Experiment.*—The following four groups of subjects were used. We will call *Group* '*S*' those subjects in whose case light was followed by a loud sound, *Group* '*P*,' light followed by air puff, *Group* '*S or P*,' light followed by sound or puff in a predetermined, unpredictable order, an equal number of each, *Group* '*S and P*,' light followed by simultaneous sound and puff.

Groups '*S*,' '*P*,' and '*S or P*' were obtained during Jan.–July 1939, consecutive subjects being assigned to each group in succession to equalize variable factors. Group '*S and P*,' not part of our original plan, was obtained in July and August, 1939.

(*h*) *Subjects.*—There was a total of 48 subjects, all college women. Group '*S*' contained 11 subjects, Group '*P*' 13, Group '*S or P*' 11, Group '*S and P*' 13.

Results and Discussion

A. Sensitization of the Lid Reflex to Light

Some of the results of this investigation bear on the problem of reflex sensitization resulting from the conditioning procedure, and will perhaps help to clarify some of the apparent contradictions of past investigations.

When two kinds of stimulus are applied to a subject during the same experimental session, their consequences may interact. This can occur even when they are not in close temporal proximity. Certain investigators have called a sensitized response to one stimulus resulting from the prior application of another, 'pseudo-conditioning.' In the case of other responses, close temporal contiguity of the stimuli is necessary for the appearance of similar phenomena. One of us (**12**) reported in 1930 that a crossed reflex to a patellar tendon blow developed in some subjects when the knee-jerk of the two legs were repeatedly elicited by closely spaced blows. This sensitization would very likely not ordinarily occur unless the stimuli were close together in time. Hilgard and Marquis (**7**, pp. 41–43, 49–50) have recently reviewed these studies. Numerous other examples of sensitization of one response as a result of prior stimulation of another could easily be listed from memory of everyday experiences.

Hilgard, 1931 (**4**), reported as one result of double stimulation by light and sound that the reflex lid closure to light increased in magnitude. Hilgard and Biel, 1937 (**5**), interpreted their results from experiments on double stimulation by light and air puff as demonstrating no sensitization of the lid closure to light. Hilgard's earlier conclusion was reached after examination of individual cases, and was based chiefly on the records of one subject; Hilgard and Biel's conclusion was reached after examination of the averages of many

subjects, and there is no indication that they were concerned with determining the extent to which the changes of individuals conformed to the group averages. Of the earlier study they say, "Since this subject had an exagerrated reflex to light to begin with, it is very possible that some pathological condition may have accounted for the results obtained" (**5**, p. 230).

Since our experiment used procedures similar both to the investigation of Hilgard and of Hilgard and Biel, it seems worthwhile to report briefly on our observations.

1. *Results*

Hilgard and Biel concluded concerning such sensitization that, "It is assuredly not the typical result of conditioning" (**5**, p. 230). Our results appear to confirm this conclusion for each condition of our experiment.

These investigators stated further, "It is a justifiable conclusion that 'alpha' conditioning [sensitization], in the eyelid situation, is a very rare phenomenon, if it may be said to occur at all" (**5**, p. 230). We doubt whether either their results or ours justify so strong a conclusion.

Considering *group averages* alone, there was a transitory increase both in average magnitude (all cases of no response were counted as zero amplitude and included in the averages) and in average frequency of the lid reflex to light in groups '*P*,' '*S or P*,' and '*S and P*,' but very little increase in group '*S*.' The period of increased average amplitude covered about the first 10 trials. The subsequent decrease in magnitude and in frequency was least in group '*S and P*,' greater in group '*P*,' and greatest in group '*S or P*.' Group '*S*' showed a slight increase in average amplitude on trial 2, thereafter a decline which reached and stayed close to zero amplitude after trial 20.

Considering the records of *individual subjects*, there were seven whose reflex lid closures to light remained greater in the last four groups of 10 trials (trials 11–20, 21–30, 31–40, 41–50) than they were in the first 10 trials, and others who showed a more transitory increase. No subject had consistently large reflexes to light comparable in amplitude to the amplitudes reached by the conditioned responses.

The average amplitude of the eyelid reflexes to light (counting all cases of no response as zero and including them in the averages) was 3.75 mm on the record during the first 10 trials for all 48 subjects.[3] It seems fair to conclude that our light stimulus was a lesser

[3] This compares to an average amplitude of 8.2 mm for the CR's of all subjects for the last 5 trials of the experiment. The corresponding frequency figures were 60 percent for the reflexes to light, 52 percent for the conditioned responses.

reflex stimulus than was used by Hilgard and Biel, since the average magnitude of the response to light in 5 unreinforced light presentations at the beginning of their experiment was 6.8 mm, obtained by a recording method giving a magnification of vertical lid movement of approximately 4 times, which is the same as ours. The average latency of the reflexes to light in our experiment was 93 ms. This figure may involve a constant error of measurement of as much as 10 ms because of the uncertainty concerning the time when the increasing incandescence of the stimulus lamp became an effective stimulus.

2. *Discussion of Lid Reflex to Light*

The results of our experiment are probably in complete agreement with those of Hilgard and Biel, but our interpretation is somewhat different from theirs. That portion of their experiment in which light and puff were separated by 400 ms (as in ours) consisted of 5 trials with light alone followed by 50 paired stimulations, and then again 10 trials with light alone. The average amplitude of the eyelid reflex to light increased when paired stimulation began, and decreased markedly when it was discontinued. (This did not occur when the interval between stimuli was 100 ms.) To present their results they calculated the alterations during the 50 reinforced trials in 10-trial averages, finding continuous decrement. It seems a reasonable guess that had they calculated trial-by-trial averages, as we did, they would have found within the first 10 trials (which they lumped together) the same average transitory increase as occurred in our experiment.

There are a number of reasons for expecting that such a more or less transitory increase would occur. (1) Peak (**10**) has shown that by appropriately controlling the set or attitude of the subject changes in magnitude of eyelid reflexes are produced. It is conceivable that a set or attitude favorable to the elicitation of the eyelid reflex to light results from paired stimulation by light and air puff, or light and combined sound and air puff. The production of such a set might be dependent upon an appropriate interval (400 ms.) between stimuli; its persistency might be expected to show wide individual variations. (2) At the beginning of an experimental session an attempt is made to put the subject at ease. The commencement of loud sounds and air puffs may be expected to interfere with such relaxation for a time, and this increased tension and alertness in turn increase reflex sensitivity (Bernstein, 1). (3) Hilgard pointed out to us (personal communication) that at the outset of the experiment the instructions to fixate the light may interfere with the reflex, and that the diminishing effectiveness of these instructions could allow

increased reflex amplitude. (4) Increasing dark adaptation may increase sensitivity.

We are not in complete agreement with the strong conclusion of Hilgard and Biel that "'alpha' conditioning, in the eyelid situation, is a very rare phenomenon, if it may be said to occur at all." The suggestive increase in their 400 ms. group, the similar results of our present experiments, the results of our analysis of records of individual subjects, the results of Hilgard and Marquis (**6**) on eyelid elevation to light in dogs,[4] as well as the theoretical aspects of the situation, seem to indicate that their conclusion was premature.

We are also in disagreement with the practice of drawing conclusions wholly from group averages. The increase in reflex amplitude shown by seven of our subjects (see above) is no less real for the fact that 35 other subjects showed decreased amplitude. One would be justified in ignoring them only if one felt quite sure either that serious errors of measurement occurred in the records (which is not the case) or if extra-experimental variables of a chance nature were operating (which cannot be demonstrated). We particularly wish to decry the practice of labeling deviant subjects as 'pathological' (**5**, p. 230).

Our excuse for so extensive a discussion of this problem of 'alpha' conditioning of the eyelid is a matter of principle. The wholly negative conclusions of Hilgard and Biel seem to us to be a result of an incorrect analysis of their data. If their conclusion is accepted, other investigators, deciding the issue is closed, may ignore or misinterpret aspects of more important data. Furthermore, one of their statements ("The results are interpreted as throwing doubt upon reflex sensitization according to conditioning principles," **5**, p. 233) may, if carelessly applied, seem to cast doubt on the accuracy or validity of previous studies of sensitization.

B. Conditioned Responses to Light

Conditioned eyelid closures (average latency about 235 ms.) developed in some subjects of each of our four experimental groups. By examination of various criteria, it seems fair to conclude that

[4] We take this opportunity to report briefly on some previously unpublished observations by Thelma Capt, F. G. Tice, and G. R. Wendt. They were concerned with the differences between dogs and monkeys, which showed eyelid elevation to light in experiments, and humans, who did not. They followed the line of reasoning that if the function of eyelid elevation to light in dogs and monkeys is to look at the light, then humans should show eyelid elevation to a light stimulus if it is placed where they must elevate the gaze in order to look at it. This was confirmed. A fixation point was placed slightly below eye level, 5 feet in front of the subject; the light stimulus appeared 1 foot above the fixation light. Five of seven subjects, uninstructed, yielded records of eyelid elevation in 34 percent of the trials. Their latency was usually in excess of one-fifth second. Voluntary lid elevations under the same conditions yielded reaction times of the order of 150 ms.

group '*S*' showed least conditioning, group '*S and P*' most, while groups '*P*' and '*S or P*' showed an intermediate amount, but were not reliably different from one another.

1. *Results*

The Table below shows for each of the four experimental groups (1) the average number, per individual, of the conditioned responses in the series of 50 trials, (2) their average recorded amplitude, counting all cases where the CR was absent as zero magnitude, and including them in the averages, (3) the percentage of subjects whose records contained 5 CR's or more, and (4) the percentage who gave as many as 3 CR's in 5 consecutive trials. By each of these criteria the effectiveness of the different conditions was as stated above.

	Reinforcement by			
	Sound	Puff	Sound or Puff	Sound and Puff
Average number of CR's in 50 trials....	5.5	16	16	28.8
Average amplitude of CR's in 50 trials	0.3 mm	0.8 mm	1.0 mm	2.1 mm
Percent of subjects with 5 or more CR's.....	36	69	73	100
Percent of subjects who reached criterion of 3 CR's in 5 trials.....	27	69	64	100

Reinforcement by loud sound (Group '*S*') was a very ineffective procedure, considering group performance. Only five of the eleven subjects gave two or more CR's; only four reached a criterion of 3 out of 5, or total of five CR's. One subject in this group, however, developed a very good CR, giving 22 responses in the last 23 trials, 16 of them above 25 mm in amplitude.

Reinforcement by simultaneous sound and puff (Group '*S and P*') was, by contrast, a very effective procedure for the development of CR's. Seven of the thirteen subjects gave at least 10 consecutive CR's. Five others gave 8 out of 10 or better. The remaining subject gave a total of 11 CR's.

Reinforcement by puff alone (Group '*P*') or by sound and puff alternated in an unpredictable order (Group '*S or P*') were intermediate in effectiveness. Comparing the two, their effectiveness seemed about equal, except for the possibility that the CR may have developed somewhat sooner in Group '*S or P*' than in group '*P*.'

Figures 1 and 2 are graphic representations of the progressive changes in average amplitude and average frequency of the condi-

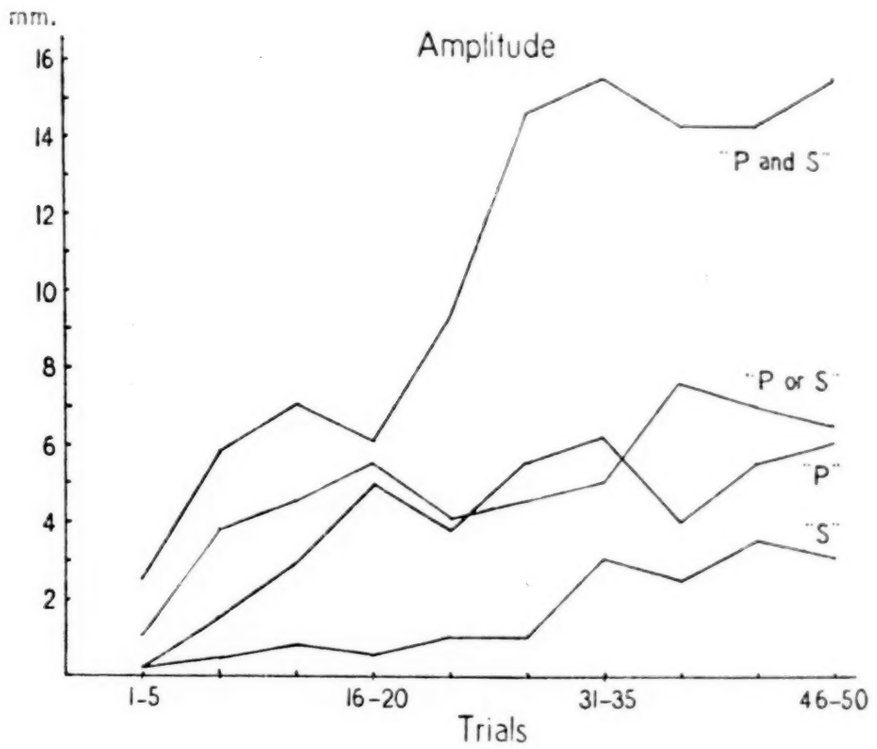

FIG. 1. Acquisition curves of conditioned eyelid closures according to amplitude. On the vertical coördinate are shown the average recorded eyelid closures. (All cases of no response are recorded as zero and included in the averages.) On the horizontal coördinate are successive groups of five trials (1–5, 6–10, 11–15, etc.).

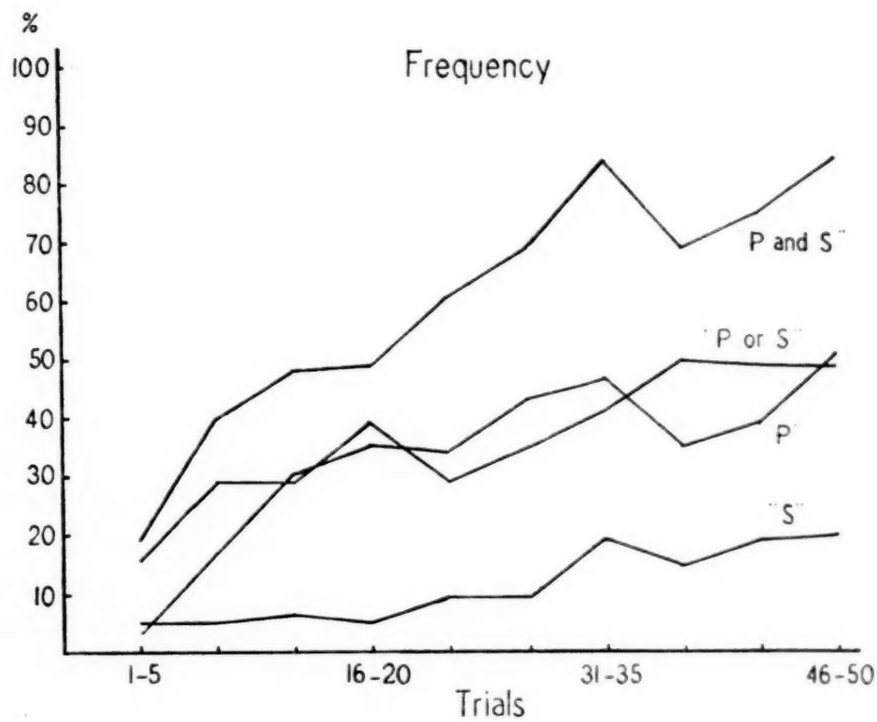

FIG. 2. Acquisition curves of conditioned eyelid closures according to frequency. On the vertical coördinate are shown the percentage of trials in which a recordable conditioned eyelid closure occurred; on the horizontal coördinate, successive groups of five trials.

tioned responses of each group during the experiment. They bear out the conclusions that reinforcement by sound is least effective, by simultaneous sound and puff most effective, and by puff or alternating sound and puff intermediate in effectiveness. They also indicate that in Groups '*S or P*' and '*S and P*' conditioned responses appeared within the first five trials. (There were also one CR in group '*P*' and two in group '*S*' within the first five trials.)

2. *Discussion of Conditioned Responses*

One's interest in this experiment may go in two directions. First, one may be interested in the extent to which the results conform to, or extend, the so-called 'principles' or 'laws' of conditioning or behavior. In this connection it may be said that our results are consistent with those of previous investigations, in illustrating the role of such factors as intensity of stimulus, heterogeneity of stimuli, dependence of rate of conditioning on magnitude of unconditioned response, etc. Second, one may be interested in the analytical problem: how, or by what mechanism, these results are produced. An answer to this second question must, in this instance, be chiefly speculative, for the experiments were not designed to answer that type of question. In the following discussion both directions of interest will be taken up together.

There has been no previous comparison within the same experiment (and therefore under really comparable conditions) of the relative effectiveness of conditioning based on loud sound and conditioning based on puff of air. Our results show the latter to be by far the more effective for conditioning of eyelid closure.[5] This relative ineffectiveness of sound as compared to air puff is probably, *in part*, the result of the rapid habituation of the response to sound. At first startling and annoying, the sound later seems much less so to most subjects, and there is an accompanying reduction in amplitude of reflex closures. With the lesser amplitude one may, according to previous investigations, expect less conditioning. The response to air puff, by contrast, shows relatively little habituation. (One needs, of course, to go behind the mere statement that the response to sound undergoes more habituation than the response to puff, to ask the reason for this difference. The consequence of failing to respond to air puff is conjunctival stimulation. Thus, the set to respond to puff is reinforced whenever lid closure fails to appear. Failure to respond

[5] The sound we used was about as loud as is feasible. Some subjects will tolerate a louder sound, but for most subjects our sound was at first unpleasantly loud. A louder sound would probably be feasible only if a series of sounds of increasing intensity, allowing for habituation, were used.

to sound will be followed by no such reinforcing consequence, hence it shows more habituation.)

It was indicated that the '*S or P*' group may have shown somewhat quicker conditioning than group '*P*.' If this is true, the explanation probably lies in the tension set up by the startling sound stimulus. The subjects' overt emotional reactions to puff were slight, but to the first few sounds they were often considerable.

The great effectiveness of simultaneous stimulation by loud sound and air puff is probably the result of a combination of the above factors. Perhaps the sound produces an initial tension, adaptation to the combination is relatively slight, and the unconditioned response retains a large magnitude.

Our original interest in this experiment was in the effects of heterogeneous reinforcement as compared with homogeneous reinforcement. On the analogy of experiments on mental and physical work, one might expect heterogeneous reinforcement at the end of 50 trials to have produced more conditioning than 25 trials of sound or 25 trials of puff reinforcement could account for. At the time we planned this experiment, the results of Brogden (**2**), Cole (**3**), and Humphreys (**9**) on reinforcement interspersed with non-reinforced trials had not yet been published. Humphreys, studying the conditioned eyelid response to light based on puff by a procedure similar to ours, found no significant difference in rate of acquisition when every light was reinforced as compared with a random alternation of reinforcement and non-reinforcement. We found no significant differences between conditioning to 100 percent reinforcement by puff and random alternation of reinforcement by sound and by puff. We cannot, therefore, decide whether our results should be interpreted in the light of previous studies of heterogeneity of stimulation, or in the light of these studies of intermittent reinforcement.

Summary

This study was concerned with a comparison of conditioning of eyelid closure of humans under four conditions of reinforcement: (1) a light stimulus followed after 0.4 second by a loud sound, (2) light followed by a puff of air to the eye, (3) light followed either by a loud sound or an air puff in an unpredictable order, (4) light followed by simultaneous loud sound and air puff. A separate group of subjects was used for each condition of reinforcement.

The results of the experiment indicated that the loud sound was the least effective for the development of conditioned eyelid closures; simultaneous stimulation by sound and air puff was most effective; while puff alone and random alternation of sound or puff were intermediate in effectiveness and approximately equivalent to one another.

Sensitization of the lid reflex to light took place as a result of the conditioning procedures. Judged by group averages, this sensitization lasted for about the first 10 trials; judged by individual records, however, 15 percent of the subjects showed slightly increased reflex closures for the entire period of 50 trials.

(Manuscript received June 13, 1941)

References

1. Bernstein, A. L., Temporal factors in the formation of conditioned eyelid reactions in human subjects, *J. gen. Psychol.*, 1934, **10**, 173–197.
2. Brogden, W. J., The effect of frequency of reinforcement upon the level of conditioning, *J. exp. Psychol.*, 1939, **24**, 419–431.
3. Cole, L. E., A comparison of the factors of practice and knowledge of experimental procedure in conditioning the eyelid response of human subjects, *J. gen. Psychol.*, 1939, **20**, 349–373.
4. Hilgard, E. R., Conditioned eyelid reactions to a light stimulus based on the reflex wink to sound, *Psychol. Monogr.*, 1931, **41**, No. 184, 50 pp.
5. Hilgard, E. R., and Biel, W. C., Reflex sensitization and conditioning of eyelid responses at intervals near simultaneity, *J. gen. Psychol.*, 1937, **16**, 223–234.
6. Hilgard, E. R., and Marquis, D. G., Conditioned eyelid responses in monkeys, with comparison of dog, monkey, and man, *Psychol. Monogr.*, 1936, **47**, No. 212, 186–198.
7. Hilgard, E. R., and Marquis, D. G., *Conditioning and learning*, New York: Appleton-Century Co., 1940.
8. Hovland, C. I., 'Inhibition of reinforcement' and phenomena of experimental extinction, *Proc. nat. Acad. Sci.*, Wash., 1936, **22**, 430–433.
9. Humphreys, L. G., The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions, *J. exp. Psychol.*, 1939, **25**, 294–301.
10. Peak, H., Modification of the lid reflex by voluntarily induced sets, *Psychol. Monogr.*, 1931, **42**, No. 188, 68 pp.
11. Robinson, E. S., Work of the integrated organism, *In* Murchison, C., *A handbook of general experimental psychology*, Worcester, Mass., Clark Univ. Press, 1934, 571–650.
12. Wendt, G. R., An analytical study of the conditioned knee-jerk, *Arch. Psychol.*, N. Y., 1930, **19**, No. 123, 97 pp.
13. Wendt, G. R., Methods of recording action, *Arch. Psychol.*, N. Y., **32**, No. 228, 83 pp.

THE FEATURES OF THE OPTIC-NERVE DISCHARGE UNDERLYING RECURRENT VISION *

BY S. HOWARD BARTLEY

Laboratory of Neurophysiology, Washington University School of Medicine, Saint Louis, Missouri

The visual experience elicited by a single short flash of light is not always that of a single short-lived sensation, but may be a variable series of pulsations depending upon flash duration and intensity and upon the intimacy of observation. Attention was first called to this sensory complexity by Young in 1872, who termed it *recurrent vision*, which has since been studied by a host of investigators.

If the pulses are spread out in space as is the case when the stimulus is an illuminated radial slit in a revolving disc, the phases may be seen as a series of light and dark bands. This procedure is not equivalent in every respect to the use of a stationary stimulus although, with due allowance, certain comparisons can be made between the results and those under conditions such as used in the present investigation. For some decades now, the sensory outcome has been divided into the following eight phases, counting the dark intervals as well as the bright pulses. (Dittler and Eisenmeier, **6**.)

(1) The primary image, the first bright flare which is the most prominent effect of the stimulus, though very transient. (2) A short dark interval lasting on the average about 40 ms and perhaps involving a faint negative phase. (3) A positive after-image, lasting about 50 ms, known as the first positive after-image, or Hering's image. (4) A second dark interval, the first negative after-image, lasting about 160 ms. (If only darkness were involved in the designation of negativity, it would logically be the second negative after-image.) (5) The second positive after-image, long called the Purkinje after-image, the secondary image, satellite, or Bidwell's ghost, appearing between 160 and 200 ms after the primary image. (6) A long dark interval, the second negative after-image. (7) The third positive, though faint, after-image, the tertiary image, or Hess's after-image. (8) A dark interval slowly shifting into an extended weak negative phase known as the third negative after-image.

This classification is not fully similar to certain others, one of which counts several phases (1, 3 and 4) as one. Be this as it may, the above classification serves to show that the sensory result is much

* This work was done under a Grant-in-Aid from the Rockefeller Foundation and from the Committee on Benevolences of the Scottish Rite Masons for Research in Dementia Precox.

different than might be anticipated from the simple character of the stimulus.

Notwithstanding the relative obscurity of some of the phases of the complete series, it is extremely easy under relatively casual conditions to obtain two prominent subjective flashes from a single pulse of light. The second, though quite well separated from the first, does not essentially differ from it and therefore is not likely to be taken for an after-image. Nevertheless identification with the second positive, or the Purkinje image, in the classification just given, has been made, as will be seen later.

In addition to the two flashes, a train of less prominent pulsations is sometimes observed under other conditions. The precise stimulus intensity and duration values required for the production or avoidance of these was not measured, although the conditions producing the latter pulsations are close to those producing the experience we are mainly interested in. Hence the phenomena have a place in the present analysis. As a part of the recognition of this, typical retinograms and optic-nerve discharge records of repetitive bursts which were often encountered in attempting to find the conditions responsible for the two-flash experience, are presented.

The purpose of the present investigation has been to compare the sensory findings with the optic-nerve discharge, the neurophysiological result of retinal activity, recently recorded in this laboratory. (Bartley and Bishop, **3**, **4**, Bishop and Bartley, **5**.) The latter activity is recorded as a picture of electrical potential representing the composite activity of the neural elements involved during each instant from the very onset of response till it is completed. It has been found that such a record contains not a smooth flow of impulses, but instead, various groupings of them and it can be assumed that these have significance for the final sensory outcome in the animal, since the cortex follows them rather precisely. If so, it might be supposed that bursts of activity in the nerve would underlie sensory pulsations, and that the record of the optic-nerve discharge and the sensory reports of human subjects would bear some resemblance under essentially similar conditions. If this were to be substantiated, it would mark a decided step in advance in the physiological study of vision, especially since some few things are known about the retinal response and since the means now at the command of physiologists promise to extend the analysis considerably further.

Method

The method consisted in comparing human observations under stimulation by short flashes of light, with records of the optic-nerve discharge of the rabbit under similar stimulation. Though the absolute intensity and duration values for the two types of experimentation were not identical

this did not logically interfere with the intended correlations since neither the species nor the conditions of the subject were equivalent. As contrasted to the fully normal humans, the rabbit, prior to recording, had been subjected to anesthesia. The essential features required to validate the comparison consisted in ascertaining the range of intensity and duration in which the dual sensory impressions were obtained, and in determining whether the necessary trait of the optic-nerve discharge became prominent only during a range located in a comparable part of the total intensity and duration scale.

To perform the human experiments an optical arrangement providing a uniformly illuminated field of any desired visual angle up to 45° was employed. An Abney episcotister was used to give the short flashes. Since the open sector on the episcotister could be varied while the disc was in motion, the duration of the flash could be progressively varied till either the minimal or maximal duration was reached without interruption to the observer. Such an apparatus has been previously employed. (Bartley 1, and in later papers.)

The patch of light selected as stimulus was 5 inches square and was viewed from a distance of 18 inches.

The description of the technique of obtaining the records from the rabbit optic nerve is given in detail elsewhere. (Bishop and Bartley, 5; Bartley, 2.)

Results

Sensory Findings.—The sensory study was confined to determining the conditions for producing the impression of two flashes from a single stimulus. It was found that a pulse of light could be either too weak or too short to produce the sensory impression of two flashes. Likewise, it could be too intense or too long. The first task then consisted in determining these limits and thus marking out an area within which the desired results would occur. The following details and the graph in Fig. 1 constitute the findings.

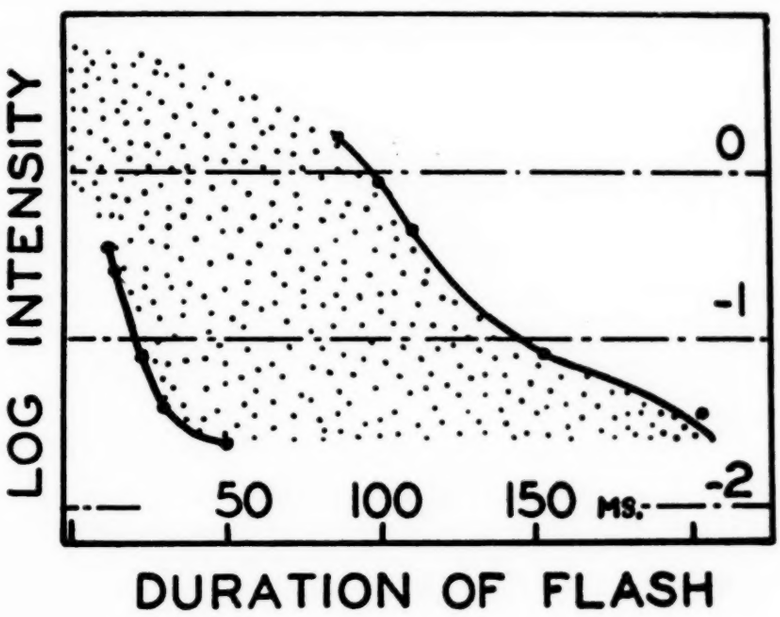

Fig. 1. A graph showing the conditions under which a single short photic stimulus gives rise to the sensation of a single flash and the conditions under which it produces two flashes. The stippled area indicates the latter, the lines bounding it show the extent to which actual measurements were made in the present experiments. The intensity of the light is in terms of candles per square foot.

With an intensity of the stimulus pulse of about 1.9 c/ft^2, the double impression exists when the pulse is at least as short as 1 ms, and continues until stimulus duration is extended to about 92 ms.

If intensity is reduced to .85 c/ft^2, the double impression is lost at 106 ms; if reduced to .66 c/ft^2, it is lost at about 112 c/ft^2; .385 c/ft^2 at 120 ms; .137 c/ft^2 at 128 ms; .0735 c/ft^2 at 155 ms, and .0343 c/ft^2 at 204 ms. Finally, due to the lowered intensity, neither flash is very strong and with stimuli longer than 0.2 second, the double impression is either becoming confused with or is being substituted by the definite *on* and *off* phases.

The flash of 1.9 c/ft^2 was too intense to be shortened below the range producing the dual impression with the apparatus at hand (1 ms). But for lower intensities measurements were obtained. A flash of .385 c/ft^2 had to be 12.5 ms long, to produce the duality; one of .0735 c/ft^2 had to be 21 ms; one of .0343 c/ft^2 had to be 25 ms; and one of .0195 c/ft^2 had to be 48 ms long. These values, like those for the upper critical durations for the double impression, are plotted in Fig. 1. The space between the two curves represents the region in which two flashes are seen by the presentation of the single light impulse, whereas the areas in both directions outside of this indicate the conditions under which only a single impression results.

To summarize, it may be said that a brief weak pulse of light produces the impression of a single short weak flash. If the stimulus is slightly more prolonged, it begins to produce the impression of two flashes in a very realistic manner, so much so that the observer would declare that two separate and closely timed stimuli were being used. If the stimulus actually used is only moderately weak, the double impression is soon lost with further increase in light intensity. If the stimulus is very weak, it does not produce the double impression until it becomes relatively long, and the duality is retained much longer as the stimulus duration is extended. There is a limit to this for when any pulse of light reaches a certain duration, the observer is able to analyze light into its three phases; *onset*, *continuation* and *cessation*. Just before the flash is separable into these three phases, the second flash of the duality has declined and disappeared.

The Optic-Nerve Findings.—Our next concern is to examine the optic-nerve discharge. Beginning with a very short weak flash of light, the response is relatively simple. Monophasically recorded, it is merely a faint deflection followed by another one several times as large. In fact, the light may be made so dim that the second wave alone can be recorded. From this point the flash may be made either longer or more intense, with the results shown in Fig. 2. If duration is extended, the size of the two waves which shall hereafter be called the *on* response is increased. As this occurs, the *on* response is fol-

lowed by a few irregular spikes or wavelets. If the flash duration is slowly increased still further, a critical point may be reached, at which the irregular sequence is substituted by a very regular train. (See Fig. 3.) Making the flash still longer soon destroys the train which is then supplanted by the former irregularity and out of this a large wave begins to loom. As this wave, hereinafter called the secondary wave, increases in size, from further extension of the flash, the prominence of the wavelets surrounding it declines. Further elongation of flash duration finally beings to reduce the height of the secondary wave and the former wavelets may again appear and take

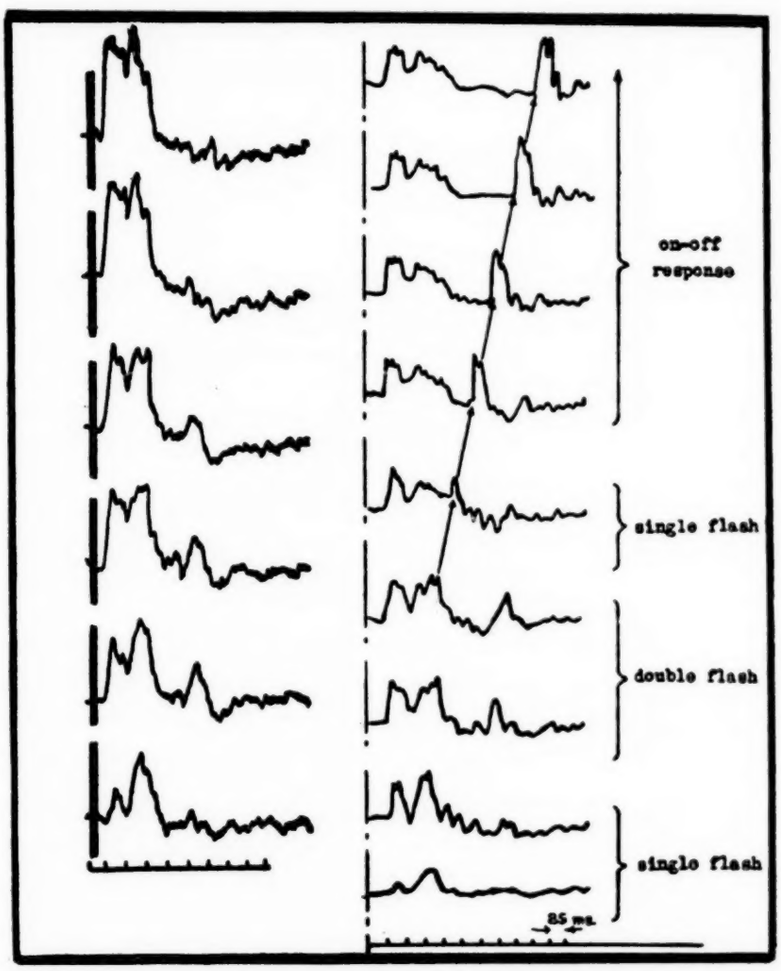

FIG. 2. Records of the optic-nerve discharge in the rabbit. The left-hand column, reading from bottom to top, shows the *on* response and the secondary wave as intensity alone is increased. The horizontal width of the vertical bars indicates the duration of the flash. The time is given below in units of 25 ms for each column. The *on* effect is the first two partly fused potentials, while the secondary wave is the much later potential that rises to a maximum as light intensity is increased and then declines. The right-hand column shows the discharge as duration alone is increased from a very short stimulus to one long enough to widely separate the *on* and *off* responses. The bottom record shows a just suprathreshold response. Reading upward, the secondary wave soon emerges and declines. Earlier in each record, the *off* response can be seen, first as a part of the second *on* response, and then as a definite spike complex finally reaching a height as great or greater than the *on* response. When the flash is long, a remnant of the secondary wave persists, still *following* the *off* response.

up positions around it. In some cases the regular train which appeared at shorter durations is almost duplicated in general characteristics. By this time the *off* response, a trace of which was previously discernible as part of the second *on* wave, has become quite definite, standing out as the first of the spikes in the train.

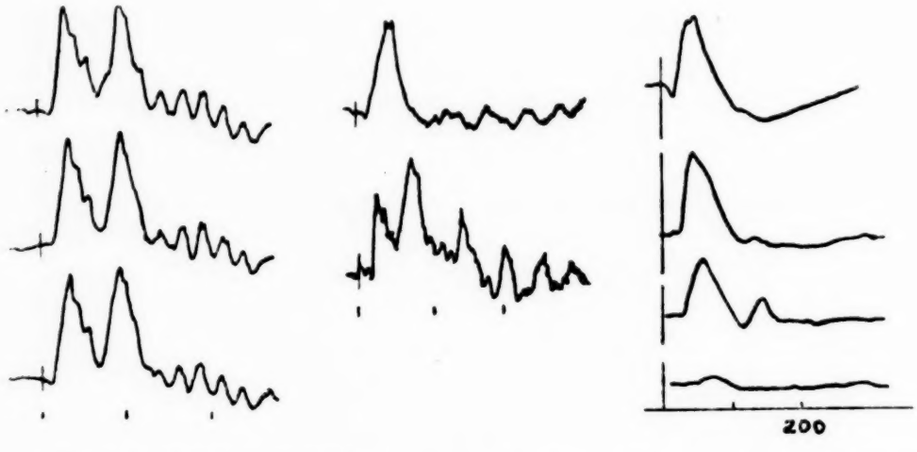

FIG. 3. Records showing repetitive response in the optic nerve and in the retina. The left-hand column shows three records at a critical duration, manifesting a train of spikes following the *on* response. These may be too small and too close together in this case to be responsible for a sensory correlate. The middle column shows a retinogram (top) containing repetitive response of the *b* wave, and an optic-nerve record manifesting a series of spikes under the same stimulus conditions. In this case, the components of the discharge are so timed as to likely be responsible for a flickering sensation. The third column contains a series of retinograms (reading from bottom to top), in which duration is lengthened step by step. Note that at a critical value the record contains a second wave which is presumably a repetition of the *b* wave (the large wave preceding it). Time in all three records is marked in units of 100 ms.

It should be noted that the secondary wave follows the *off* response, and by the time the former has become quite noticeable the latter has declined to a small fraction of its maximum height, running the chance of being overlooked unless a consistent series of records is examined. With each addition to flash duration, the *off* response shifts to a later place in the record, soon becoming large enough to usurp attention, the secondary wave shifting also (though to a lesser degree), and always following the *off* response.

If flash intensity rather than duration is increased, the secondary wave likewise appears, the uniform train of wavelets or spikes occasionally not preceding it, for which failure experimentation may be at fault in having passed through the critical intensity by too large steps. As intensity is increased, the size of the *on* response increases but relatively more slowly than the secondary wave. Finally the secondary wave reaches a maximum and declines. Accordingly, there is an optimum intensity for the maximum prominence of the secondary wave, and this is reached at a considerably lower stimulus intensity than for the maximum *on* response. By the time that the

on response becomes very large, the secondary wave has declined to a mere remnant of its former size.

The Retinogram.—Not only the optic-nerve discharge but also the record from the eyeball (the retinogram) exhibits waves beyond the initial response potential (*b* wave), under critical conditions when the eye is confronted with a single brief pulse of light. Figure 3 contains examples of this phenomenon. In the retinogram these extra waves appear to be repetitions of the *b* wave itself, and inasmuch as they appear at intervals which are matched by spikes in the optic-nerve discharge under identical conditions, the events in the two series may eventually serve to explain each other. We know that the retinogram does not display an *off* response, and for that reason the retinogram would not seem to be the manifestation of synaptic activity in the retina, but instead represents the activity of the sense-cells. Since, however, the retinogram has been analyzed into several components each apparently somewhat independent of the other, we cannot say anything conclusive as yet. The retinogram has been introduced at this time only to suggest, in a concrete way, the eventual possibility of the further analysis of the neural components responsible for the specific items in the sensory response.

Discussion

The Requirements.—Although the presentation of data includes the sensory phenomenon of flutter or flicker produced by a single pulse of light, records of trains of spikes in the optic nerve, and repetitive responses in the retinogram, our chief interest lies in correlating the two-flash experience with events in the optic-nerve discharge. In order that the optic-nerve discharge and the sensory phenomenon be shown to correlate, certain specific requirements are dictated. (1) Since the sensation in question is that of two pulses of light, the optic-nerve discharge should exhibit two bursts of activity. (2) Since the second of the two pulses appears only under a limited set of conditions, the emergence of the two bursts of activity in the nerve discharge should depend upon a similar set. (3) Since the two sensory pulses are temporally distinct, the two bursts in the nerve discharge should be quite discrete. It is to be admitted that two very closely spaced or even partially overlapping potentials could conceivably be responsible for temporally discrete sensory effects through a delay in the later part of the pathways responsible, but if so, the correlation would be much less obvious, and might have to be abandoned. (4) Since all of the distinguishable phenomena following the very first pulse of light are placed in the category of after-effects, some justification for this is to be looked for in the optic-

nerve discharge, though a failure to find it would not necessarily be injurious to the correlation. The fact that certain phenomena are classed as after-images rather than as part of the primary sensations has depended upon point of view. Since we now have the optic-nerve activity to include in our knowledge, there may be something more tangible to base the interpretation upon.

Requirements similar in principle are also dictated for other correlations between the two sets of events. For example, the sensory impression of flutter or flicker would require a series of properly spaced bursts in the nerve discharge.

The Correlation.—The comparison of the two sets of results, sensory and the neurophysiological, indicates that they have the correlative features required for crediting the major waves in the optic-nerve discharge with representing the subjective phenomena in question. Even though the optic-nerve investigation was made on the rabbit and the sensory observations on the human, the compairson is surprisingly good. As the records already presented show, the optic-nerve discharge displays a second potential which can be taken as the activity responsible for the second pulse in sensations. This second potential or secondary wave is seen under intensity and duration conditions similar in point of location in the total stimulus scale to those for the sensation. The secondary wave follows the *on* effect at a distance much greater than the separation of the two potentials which constitute the *on* response. In fact, the secondary wave arises about 200 ms after the beginning of the *on* response.

All of the requirements seem to indicate it is the progenitor of the second sensory pulse, and the position of the wave with regard to the beginning of the *on* response tends to identify it with the Purkinje after-image (Bidwell's ghost) in the conventional psychological analysis. Once granting this identiry, our inquiry leads us to consider whether or not the secondary wave and the second subjective flash are phenomena in their respective series which could rightly be called after-effects.

It appears the earlier sense-psychologists reasoned that sensation should match the stimulus in quantitative features; hence, a single short stimulus, a single short sensation. Since they found no response was that limited, they had to think of the additional phenomena beyond their expectations as *extra-* or *after*-effects, and so they called them after-images, rather than sensations at all. This was a simple view based only on sensory experimentation. When a flash of light was the stimulus, all that followed in sensation after a short single pulse of brightness was an after-effect, each component of which should be called an after-image, or after-sensation. We are now in a better position to offer an opinion, since we have been able to

examine the physiological basis of sensation. In this connection, examination of the optic-nerve discharge has revealed that it is characterized by an *on* effect and an *off* effect. These are identifiable under a wider range of conditions, and it would seem then that any event in the physiological record that precedes the *off* effect should be classed as a part of the sensation proper, while an event following it could be classed as an after-effect.

The *off* effect would thus serve as the boundary between what is response and what is *after*-response. In all cases in which events in the optic-nerve discharge can be identified as the progenitors of sensory phenomena, the distinction can be directly applied. All sensory phenomena originated by activity preceding the *off* effect are part of sensation proper, and those originated by activity following the *off* effect are to be classed as *after*-sensations, when a distinction is desired.

In light of this view, a definite conclusion can be reached with reference to the secondary wave and its sensory consequence, the second bright pulse of light. When the secondary wave first comes into prominence, the *off* effect is still a part of the second potential of the *on* effect. As the stimulus flash is made longer and longer, the *off* effect becomes definitely identifiable and moves later and later. From the very start, considerable distance separates the *off* effect and the secondary wave, the secondary wave arising *after* the *off* effect. As the *off* effect reaches prominence, the secondary wave has greatly declined in size, leaving a remainder little more than identifiable in the record. As the *off* effect becomes later, owing to prolongation of the flash, the remnant of the secondary wave, as long as it still exists, shifts also, and still follows the *off* response. It can be seen then that the second pulse of light on the sensory level may well be called an *after* effect, thereby justifying the older workers' classification.

It is also possible to raise the question as to whether the secondary wave might not be a part of the *off* effect. This can be answered in the negative, for when the *off* effect is smallest, the secondary wave is reaching maximum and is well separated from it. When the *off* effect reaches maturity, the secondary wave has declined and, though still following it, occupies a different temporal relation to it.

A further suggestion can be made; namely, that the *off* effect, which has by this time become identifiable, plays a part in producing the dark interval between the two sensory pulses. Increases in flash duration, from this point on, enhance the *off* effect and diminish the secondary wave. Just so, do increases in stimulus duration enhance the experience of the darkness and diminish the weight of the second pulse till it is insignificant.

The secondary wave is perhaps not so large in our rabbit records as we should expect it to be to represent so strong a pulsation as is observed in the human sensory experiments, but this may be the unavoidable result from the differences in the states of the two species, determined by the methods of experimentation, or its size may also represent a real difference which would exist in spite of any difference in experimental procedures.

The present study is a demonstration that some after-images can be peripherally originated in the same way that primary sensations are, though it admits a probable distinction between such after-images and those that occur much longer after the primary sensation has ceased. In this connection, we have no optic-nerve records showing a long drawn-out discharge lasting for seconds after the cessation of illumination. Nevertheless the discharge necessary for their production might be so weak as to escape detection by present physiological methods.

We have in the present an example in which physiological information of a kind unobtainable from the human subject reveals a fairly close correlation with human sensory experience. Though this correlation as it stands is not between the physiology of the human eye and that of human sensation, it could hardly be thought that a poorer correlation would be found were it possible to employ the human eye instead of that of the rabbit. The findings thus present the implication, namely, that even though the human and the rabbit make widely dissimilar uses of vision, the afferent outputs (sensory messages) of the two eyes are remarkably similar.

Summary

The intensity and duration of a short flash of light which evokes the subjective impression of two was measured.

The optic-nerve discharge under like stimulus conditions was also recorded (in the rabbit) and the outcomes in the two cases were compared.

The correlation was remarkably close, not only between these two phenomena in the two spheres, but also between other features such as flicker and the train of spikes in the nerve record.

Retinogram records which manifested a sequence appearing to be a repetition of the *b* wave were presented to throw light on the connection between the retinogram and the optic-nerve discharge.

It was shown that the second subjective pulse of the two evoked by the single flash could be called an after-effect (after-image) owing to its emergence *following* the *off* response.

The study is one of a long series of investigations in this laboratory elucidating the physiological behavior of the visual pathway and connecting the events in it with the sensory phenomena of vision.

(Manuscript received June 6, 1941)

References

1. Bartley, S. H., The relation of intensity and duration of brief retinal stimulation by light to the electrical response of the optic cortex in the rabbit, *Amer. J. Physiol.*, 1934, **108**, 387–408.
2. Bartley, S. H., *Vision: A Study of its Basis.* 1941. Van Nostrand. N. Y.
3. Bartley, S. H., and Bishop, G. H., Optic-nerve response to retinal stimulation in the rabbit, *Proc. Soc. Exp. Biol. and Med.*, 1940, **44**, 39–41.
4. Bartley, S. H., and Bishop, G. H., Some features of the optic-nerve discharge in the rabbit and cat, *J. Cell. Comp. Physiol.*, 1942 (in press).
5. Bishop, G. H., and Bartley, S. H., Activity in the optic system following stimulation by brief flashes of light, *Proc. Soc. Biol. and Med.*, 1941, **46**, 557–558.
6. Dittler, R., and Eisenmeier, J., Über das erste positive Nachbild nach kurzdauernder Reizung des Sehorganes mittels bewegter Lichtquelle, *Arch. f. d. ges. Physiol.*, 1909, **126**, 610–647.

DISTRIBUTED PRACTICE IN MOTOR LEARNING: SCORE CHANGES WITHIN AND BETWEEN DAILY SESSIONS

BY ERNEST R. HILGARD AND M. BREWSTER SMITH

Stanford University and Harvard University

Comparison of distributed and massed practice has both practical and theoretical significance. Practical interest centers in the determination of the circumstances leading to more efficient learning. Theoretical significance attaches to data testing hypotheses respecting 'reminiscence' (**2, 5**), 'maturation' (**3, 11**), 'growth' (**4, 8, 9**), as explanations of the score gains over periods of rest. These gains with rest are by no means universal; decreases in score with rest are commonly found later in learning (**8**). The limits within which gains and losses are found need to be determined with greater precision before hypotheses can be stated more appropriately, and before these hypotheses can be submitted to crucial test. This investigation was designed to provide further data on score changes within and between daily sessions by groups of subjects having different distributions of practice and rest.

PROCEDURE

Subjects practiced individually on the pursuit rotor as described by Koerth (**6**). This consists of a $\frac{3}{4}$ inch brass target revolving with the turntable of a phonograph at 1 revolution per second. The problem is to keep the tip of the hinged pointer on the target. Success is scored by means of an electro-magnetic counter which records 10 counts per revolution if the pointer is in continuous contact with the target. Since a trial is of 1-minute length the possible score is 600 per trial.

Three groups of subjects, each practicing for 4 daily 25-minute sessions, were differentiated by the number of 1-minute trials within each session. The first 2 trials, alike for all subjects, were spaced at a 1-minute interval. Following these trials, the subjects of Group A had 5 trials, each separated by 3 minutes, then a final trial after a 1-minute rest, a total of 8 trials within the session. Subjects of Group B had 1-minute rests throughout, practicing for 13 trials within the session. For the subjects of Group C the first 2 trials were followed by 15 trials separated by 20-second rest intervals, the final trial following, as in each of the other groups, a 1-minute rest. There were thus 18 trials per session for Group C. The arrangement of the trials within the daily sessions, as described above, may be somewhat more easily visualized from Table 1.

The daily sessions were all alike in arrangement, occurring on 4 consecutive days. There were 24 subjects each in Groups A and B. Subjects were alternated in the running of these groups, and the groups were kept approximately matched by initial scores. Group C was added later. By chance the initial scores of Group C averaged somewhat lower than those of Groups A and B on the trials in which all subjects were treated alike. Several additional subjects were run in Group C, bringing the total to 30. The only subjects rejected were a few who, for reasons

TABLE 1

ARRANGEMENT OF TRIALS WITHIN EACH DAILY SESSION

Minutes from Start of Practice to End of Given Trial	Consecutive 1-minute Trials		
	Group A Trials 3 to 7 Each Follow 3-minute Rests	Group B Trials 3 to 12 Each Follow 1-minute Rests	Group C Trials 3 to 17 Each Follow 20-second Rests
1:00	(1)	(1)	(1)
3:00	(2)	(2)	(2)
5:00		(3)	(3)
			(4)
7:00	(3)	(4)	(5)
9:00		(5)	(6)
			(7)
11:00	(4)	(6)	(8)
13:00		(7)	(9)
			(10)
15:00	(5)	(8)	(11)
17:00		(9)	(12)
			(13)
19:00	(6)	(10)	(14)
21:00		(11)	(15)
			(16)
23:00	(7)	(12)	(17)
25:00	(8)	(13)	(18)

beyond the control of the experimenters, failed to repeat the 4 days of experimentation. All subjects were college students.[1]

Subjects were permitted to know their scores as they proceeded. While this probably influenced results, especially in the direction of an end-spurt on the last trial within the daily session, there is no reason to suppose that it affected the several groups differentially. The subjects rested or engaged in conversation with the experimenter during the inter-trial periods.

SCORES WITHIN DISTRIBUTED AND MASSED PRACTICE

The gross findings may be summarized by reference to the practice curves plotted in Fig. 1. Mean scores for each of the groups are plotted against time elapsed from the start of practice. Since Group A had 8 trials per day, Group B had 13, and Group C had 18 trials, Group C will have had more trials than Group B at any given time late in practice, and Group B will in turn have had more trials than Group A. Inspection of Fig. 1, prior to more careful statistical analysis, suggests the following interpretations:

(1) It appears that the subjects of Group C, with the 20-second rests (massed practice), are at their greatest disadvantage in the first

[1] The experimentation was done at Stanford University within the academic years 1938–1939 and 1939–1940.

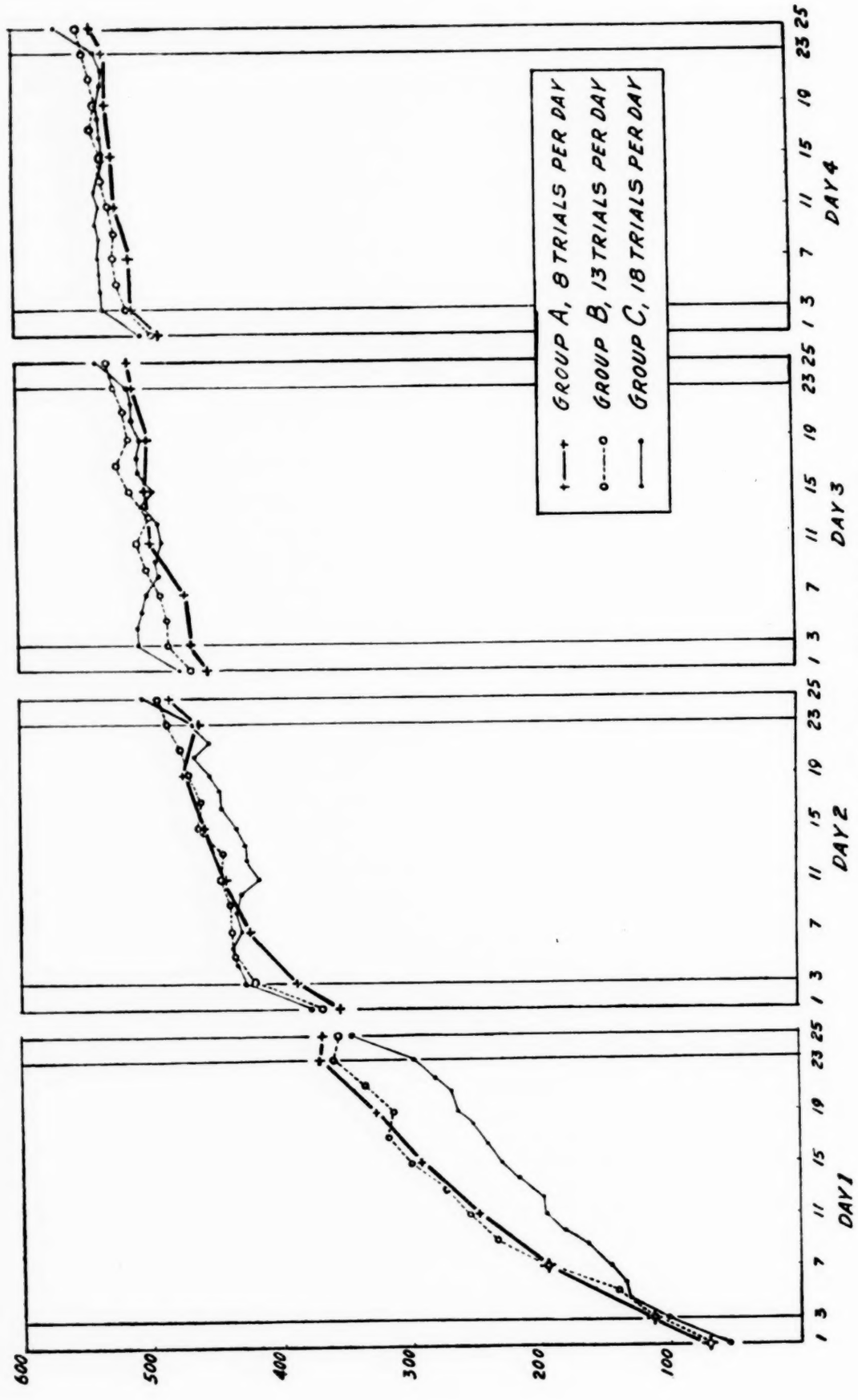

FIG. 1. Mean scores throughout practice for groups differing in the distribution of practice and rest. Trials of 1-minute duration throughout, with possible score of 600. The second and last trials of each day follow 1-minute rests. Other trials follow 3-minute rests for Group A (24 subjects), 1-minute rests for Group B (24 subjects), and 20-second rests for Group C (30 subjects). Mean scores are plotted above that

daily session, while Groups A and B score very much alike. Group C, with the most practice, makes the porest scores on this day.

(2) Scores at the end of each of the 4 daily sessions are very much alike, in spite of the differences in amount of time spent in practice. This means not only that there is an advantage to distributed practice throughout the period studied, but it suggests that there may be a 'typical' course of improvement with time, achieved within wide variations in practice.

(3) All groups show overnight losses between Days 2 and 3 and between Days 3 and 4. While these losses with rest limit to some extent the advantages of distributed practice, distribution within sessions is still advantageous at the stage of practice in which an overnight loss is found.

TABLE 2

EFFECTS OF MASSED AND DISTRIBUTED PRACTICE WITHIN DAY 1
(Trial numbers are given in parentheses)

	Minutes from Start of Practice to End of Trial	Group A (n = 24) M and σ_m	Group B (n = 24) M and σ_m	Group C (n = 30) M and σ_m
Preliminary trials	1:00	(1) 66±10.0	(1) 66±12.2	(1) 55± 7.3
	3:00	(2) 114±15.2	(2) 112±18.7	(2) 108±12.0
First trial after differential rest intervals	4:20			(3) 130±14.6
	5:00		(3) 153±22.1	
	7:00	(3) 196±20.2		
Last trial after differential rest intervals	23:00	(7) 371±19.4	(12) 362±19.3	(17) 297±16.3
Final trial after 1-minute rest	25:00	(8) 369±18.6	(13) 359±22.5	(18) 354±16.3

Since these interpretations bear upon theories of distributed practice, it is important that they be substantiated by more careful statistical analysis of the data.

The effects of distributed practice within Day 1 may be studied in relation to the scores presented in Table 2. Within the first 2 trials, during which all subjects were treated alike, mean scores did not differ significantly. A single differential rest period resulted in scores on the third trial falling in the order of the length of the rest interval, the attained score being 130 for Group C with the 20-second rest, 153 for Group B with the 1-minute rest, and 196 for Group A with the 3-minute rest. The critical ratio of the extreme difference (Group A compared with Group C) is 5.08, a clearly significant value.[2] By

FIG. 1. Mean scores throughout practice for groups differing in the distribution of practice and rest. Trials of 1-minute duration throughout, with possible score of 600. The second and last trials of each day follow 1-minute rests. Other trials follow 3-minute rests for Group A (24 subjects), 1-minute rests for Group B (24 subjects), and 20-second rests for Group C (30 subjects). Mean scores are plotted above that

[2] Since the 3 groups may be considered matched according to the second trial, Wilks' formula was applied in testing the significance of differences at later points. The relevant correlations, obtained from a pooling of all 78 subjects in accord with the null hypothesis, are +.88 between the second and third trials on Day 1, +.65 between the second trial and the last trial under conditions of differential rest (second-last trial), and +.75 between the second trial and the last trial of the same day.

the time of the last trial under conditions of differential rest, Groups A and B scored very much alike, although by this time Group A had had 7 trials, and Group B had had 12 trials. The 20-second rest continued unfavorable for Group C, however, so that its mean score differed from Group B with a critical ratio of 3.35 and from Group A with a critical ratio of 3.80. In spite of its greater number of trials, Group C's mean fell below those of the other groups. The final trial, after a 1-minute rest within each group, brought each of the groups to a common level of scoring. No significance is to be attached to the slight differences in the final means, which happen to be in an order inverse to the number of trials.

Following a single longer rest between the last 2 trials of Day 1, Group C recovered to approximate equality with Groups A and B. This suggests that the scores of Group C following 20-second rest

TABLE 3

DAILY INCREASE IN SCORE FROM FIRST TO LAST TRIAL EXPRESSED AS AMOUNT OF INCREASE PER TRIAL

	Score Increase per Trial		
	Group A 8 Trials per Day	Group B 13 Trials per Day	Group C 18 Trials per Day
Day 1	37.8	23.0	16.6
Day 2	16.4	9.4	7.1
Day 3	7.6	5.3	3.7
Day 4	6.0	4.5	3.5

periods may have been depressed by the same factors found in work-decrement experiments when insufficient rest is provided between work periods. Such an interpretation, however, does not account for the approximate *equality* of all scores within the last trial of the day in spite of the differing amounts of the period spent in practice. The decremental effects of massing of trials in Group C were also shown within the first few trials of Days 2 and 3, during which there were actual decreases in score before increases again set in.[3]

The relative uniformity of final scores after each of the daily practice sessions means that the advantages of distribution continue, since there are great inequalities in the number of trials. The differences among the groups may be expressed in terms of score increases per trial on each of the days. Such data are provided in Table 3. The decrease in gains from day to day within each group is simply an expression of the deceleration of the learning curve; the relative gains within the groups remain very similar from day to day,

[3] This feature of curves of pursuit learning after resumption of the task following long rests has been noted by Bell (1) and by Melton (7).

Group A throughout making the greatest gain per trial, Group C the least. Whatever happens within time periods as short as 20 seconds to 3 minutes continues to operate rather similarly throughout the 4 days. Trial distribution retains its advantages late as well as early despite any effect of that characteristic of 'lateness' in learning responsible for overnight losses. The analyses of Tables 2 and 3 support the inspectional interpretation of Fig. 1 in showing the advantages of distributed practice both on Day 1 and throughout the remaining days of the experiment.

Overnight Changes in Score

The initial and final mean scores of each session for each group are presented in Table 4, along with the standard deviations of the scores. While between Day 1 and Day 2 Groups B and C showed a mean gain, between all other sessions there was a mean loss over-

TABLE 4

Initial and Final Scores of Each Day
(Cumulative trial numbers in parentheses)

		Means and Standard Deviations		
		Group A $M = 24$	Group B $M = 24$	Group C $M = 30$
Day 1.	Initial	(1) 66±48	(1) 66± 59	(1) 55±39
	Final	(8) 369±89	(13) 359±108	(18) 354±88
Day 2.	Initial	(9) 354±66	(14) 370± 74	(19) 377±68
	Final	(16) 485±51	(26) 492± 66	(36) 505±47
Day 3.	Initial	(17) 455±46	(27) 463± 68	(37) 473±55
	Final	(24) 516±34	(39) 532± 41	(54) 539±40
Day 4.	Initial	(25) 492±45	(40) 493± 63	(55) 502±45
	Final	(32) 540±29	(52) 551± 30	(72) 565±29

night for all groups. Gains were shown between Day 1 and Day 2 by the following numbers of individual subjects: 10 of 24 in Group A, 13 of 24 in Group B, 20 of 30 in Group C, or a total of 43 of the 78 subjects. Between Day 2 and Day 3, only 13 subjects showed an overnight gain, while between Day 3 and Day 4 the gain was shown by 12 subjects. Thus between the first 2 days, 55 percent of the 78 subjects gained, while 45 percent lost (or remained unchanged) and between the last 2 days only 15 percent gained, while 85 percent lost. The mean results are therefore rather typical of the behavior of scores for the individual.

Just what is happening overnight is difficult to infer from the data, and a separate experimental analysis is probably required. Between

Day 1 and Day 2 there is a significant decrease in the standard deviation of the scores. For the 78 subjects considered as one group, the final score on Day 1 averages 359, with a standard deviation of 98, while the score on the first trial of the next day is 369, with a standard deviation of 71. The change in standard deviations may be calculated by the conventional formula [4] to be 4.1 times its standard error, sufficient to give confidence that the change has not come about by chance. Since the mean score changed so little overnight, it may be stated as an empirical observation that the subjects behaved more alike at the beginning of Day 2 than they did at the end of Day 1.

The contrary empirical situation exists between each of the other days; that is, associated with the overnight decrease in score there is in every case but one an increase in the standard deviation, so that subjects appear to be behaving less alike after the night's rest. The meaning of this generalization is not clear, however, in view of the relationship between standard deviation and mean as scores reach higher values. Because of the constriction of scores at the top, late in learning it is found that the higher the mean score the lower the standard deviation. Hence with appreciable overnight loss in mean, increase in standard deviation is to be expected, even though no new causal factor enters to distort individual scores after rest. Statistical correction of the relationship of standard deviation to size of score by ordinary methods involves assumptions which cannot be fully supported, so that such corrections are not attempted at this time.

Effect of Single Lengthened or Shortened Interval between Trials at Beginning and End of Each Session

If interpolated time has an important bearing upon changes in score, single shortened or lengthened intervals should affect gains (or losses) on successive trials. In the arrangement of the experiment, the first 2 trials of the day and the last 2 were in every case separated by 1-minute rests. Following the second trial and preceding the second-last trial of each day the intervals were 3 minutes for Group A, 1 minute for Group B, and 20 seconds for Group C. Thus at the beginning and end of each day there are sets of score changes over increased and decreased intervals (for Groups A and C) which may be compared with the changes over a uniform interval in the case of Group B. The changes are summarized in Table 5.

Consider first the interval changes involving the first 3 trials of each day. On Day 1, lengthening the interval increases the gains of Group A, shortening the interval decreases the gains of Group C,

[4] Taking into account the correlation of +.77 between scores at the end of Day 1 and the beginning of Day 2.

as compared with Group B which, with uniform interval, serves as a control. On the other days the negative values for Group B reflect the deceleration of gains following the rapid recovery from overnight losses between the first 2 trials of the day. Except for Group A on Day 4, all other comparisons at the beginning of each day are harmonious with results for Day 1, Group A showing some advantage due to the longer rest, Group C showing some disadvantage relative to Group B. In general, the results show that a single longer rest after the second trial results in a relative gain, while a single

TABLE 5

EFFECT OF SINGLE LENGTHENED OR SHORTENED INTERVAL BETWEEN TRIALS

	Score Gain over the Second Interval Minus the Gain over the First Interval					
	Interval Changes Involving First 3 Trials of Each Day			Interval Changes Involving Last 3 Trials of Each Day		
	Group A 1 min. to 3 min.	Group B 1 min. to 1 min.	Group C 1 min. to 20 sec.	Group A 3 min. to 1 min.	Group B 1 min. to 1 min.	Group C 20 sec. 1 min.
Day 1	+34	−5	−31	−44	−27	+41
Day 2	+5	−33	−42	+28	−6	+30
Day 3	−9	−21	−34	−5	−1	+24
Day 4	−17	−14	−27	+12	−4	+29
Mean, Days 2–4	−7.0	−22.7	−34.3	+11.7	−3.3	+27.7

shortened rest results in a relative loss. The amount of the difference varies as practice continues. This is especially evident in the comparison of Groups A and B, in which the daily advantages for Group A over Group B are +39, +38, +12, and −3. It appears that the advantage is decreasing, so that the one discordant value, on the final day, is coherent with the trend. Had the experiment been continued further, Group A might have continued to lose any advantage from the single lengthened rest following the second trial, the lengthened rest becoming actually detrimental.

The interval changes involving the last 3 trials of each day may next be considered. On every day, lengthening the interval for Group C from 20 seconds to 1 minute produces a disproportionate gain relative to Group B, the control comparison. On Day 1, shortening the interval for Group A from 3 minutes to 1 minute causes a loss, but on the second and fourth days there is a gain relative to Group B. The trend of the difference is shown by the following daily advantages of Group A over Group B (beginning with the disadvantage on Day 1): −18, +34, −4, +16. The average effect is

towards an advantage for the shorter interval, a trend coherent with the results late in practice for the beginning of the day.

To summarize these comparisons, it may be stated, first, that a 1-minute rest is invariably more favorable than a 20-second rest in producing increases in score at both the beginning and end of the day both early and late in practice; second, a single 3-minute rest, favorable to a 1-minute rest at both the beginning and end of the first day, loses its advantage later in learning. While it appears that shortening of rests might be advantageous later in learning, the shortening must not be carried to extremes, for it is evident that in this situation continuous practice (approximated by the 20-second rest intervals) is always disadvantageous.

Discussion

The interrelationships discussed in the preceding sections bear upon the theories which have been proposed to account for the results of distributed practice in motor learning.

The general disadvantage of the group with 20-second rests (Group C), particularly on the first day, gives some support for interference or work-decrement theories (**4**, **5**). The recovery with short rests at the end of each day is in line with such interpretations. The decreases in score within the first few trials of later days also give evidence of some sort of interfering process against which gains have to be made if they are to occur at all. These results are coherent with Travis' finding of the unfavorable effects of continuous practice (**10**), and with the diminishing scores for massed practice after rest as found by Melton (**7**) and Bell (**1**).

Since scores at the end of each day were very similar for each of the groups in spite of the accumulating differences in total trials, some weight must be given to theories which emphasize a characteristic rate of improvement within wide ranges of practice (**3**, **11**).

While the majority of subjects gained overnight between the first and second day, most of the subjects lost overnight between later days. These overnight losses agree with the empirical findings of Snoddy in experiments with mirror-tracing (**8**), and confirm a distinction between changes early and late in pursuit learning reported by Bell (**1**). The distinction between 'earliness' and 'lateness' enters also in the interpretation of changes introduced by single lengthened or shortened rests. A lengthened rest early in learning always produced a disproportionate gain, while a single shortened rest was always detrimental. Late in learning, however, the relative advantage changed, so that a single rest shortened from 3 minutes to 1 minute was beneficial in some of the comparisons. A rest shortened

to 20 seconds was always unfavorable. Except for the fact of a lower limit of favorable rest interval, these findings are also in agreement with Snoddy's assertions with respect to interpolating time early and late (**8**).

The progressive changes taking place as practice continues raise other questions. For example, in the case of Group C following overnight losses there are rapid gains for a few trials before the decreasing scores set in; then there is an upturn and final scores are similar on each day to the scores of Groups A and B. Groups A and B likewise show rapid gains in the early trials of each day, so that overnight losses are quickly overcome. Whether these rapid gains at the beginning of each day are a relearning after forgetting, or simply a 'warming-up' effect cannot be settled without further experimental analysis. The problem of changes in variability with practice has also been raised. The empirical findings that variability at first increases and then decreases can have no meaning until some satisfactory correction is applied for units of scoring early and late in practice. Such a correction is needed also before satisfactory interpretations of overnight changes can be made.

The agreements among the empirical findings from diverse experiments in perceptual-motor learning are impressive. These tasks appear to be fruitful situations in which to study a number of significant generalizations about learning. It is to be hoped that enough of the studies will be done under conditions sufficiently comparable that more extensive quantitative systematization may presently be attempted.

Summary

The 78 college students who served as subjects practiced individually on a Koerth pursuit rotor for 4 daily sessions of 25 minutes each. Trials were uniformly of 1-minute length, but rests differed for the several groups so that the number of trials per day were unlike: Group A (24 subjects) had 8 trials per day, Group B (24 subjects) had 13 trials, and Group C (30 subjects) had 18 trials. The advantages of distributed practice were evident throughout, being especially marked on the first day. In spite of the widely differing number of trials, all groups scored very much alike at the end of each daily practice session.

While 55 percent of the 78 subjects gained overnight between the first and second day, overnight losses were found for 85 percent of the subjects between consecutive days thereafter. These overnight changes bore little relationship to the distribution of trials within each period.

Scores following changes in single rest intervals between trials early and late in each day showed a 20-second interval to be unfavorable to a 1-minute rest in all comparisons. Early in learning a 3-minute rest was always more favorable than a 1-minute rest, but later in learning a change from a 3-minute rest to a single 1-minute rest occasionally brought a relative increase in score.

The results are discussed in relation to theories of distributed practice, particularly those stressing interference and those stressing growth.

(Manuscript received June 23, 1941)

References

1. Bell, H. M., Effects of early and late rests in motor learning as a test of Snoddy's hypothesis, Unpublished Ph.D. dissertation, Stanford University, 1941.
2. Buxton, C. E., Retroaction and gains in motor learning: III. Evaluation of results, *J. gen. Psychol.*, 1940, **22**, 309–320.
3. Doré, L. R., and Hilgard, E. R., Spaced practice and the maturation hypothesis, *J. Psychol.*, 1937, **4**, 245–259.
4. Doré, L. R., and Hilgard, E. R., Spaced practice as a test of Snoddy's two processes in mental growth, *J. exp. Psychol.*, 1938, **23**, 359–374.
5. Hovland, C. I., Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice, *J. exp. Psychol.*, 1938, **22**, 201–224.
6. Koerth. W., A pursuit apparatus; eye-hand coordination, *Psychol. Monogr.*, 1922, **31**, No. 140, 288–292.
7. Melton, A. W., The effect of rest pauses on the acquisition of the pursuit-meter habit, Paper delivered at Midwestern Psychological Association, Athens, Ohio, April 12, 1941.
8. Snoddy, G. S., *Evidence for two opposed processes in mental growth*, Lancaster, Pa.: Science Press, 1935, pp. 103.
9. Snoddy, G. S., A reply to Doré and Hilgard, *J. exp. Psychol.*, 1938, **23**, 375–383.
10. Travis, R. C., Practice and rest periods in motor learning, *J. Psychol.*, 1936, **3**, 183–187.
11. Wheeler, R. H., and Perkins, F. T., *Principles of mental development*, New York: Crowell, 1932, pp. xxvi + 529.

A FURTHER STUDY OF THE FUNCTION OF REWARD

BY HANS WALLACH

Swarthmore College

AND

MARY HENLE

University of Delaware

In a previous paper[1] we reported experiments which were designed to analyze the Law of Effect. The experiments now to be described present further evidence for our previous observations.

The term 'Law of Effect' has been used in two different meanings: it has been used to refer to the general observation that reward plays an important role in learning, or it has meant the very specific hypothesis which Thorndike advanced in explanation of this observation. According to Thorndike's hypothesis, a satisfying state of affairs which follows an organism's reaction to a given stimulus has the automatic effect of strengthening the connection between the stimulus and the reaction. This hypothesis thus claims that the effect of reward on memory is direct and independent of all other factors, for example of the nature of the subject's motivation (provided that the motivation does not interfere with the character of the reward). This is the simplest possible hypothesis to explain the effect of reward. It is this directness and independence of larger contexts which is the concern of our investigation.

Thorndike has performed one type of experiment which sets up particularly well those elements which he considers to constitute the basic scheme of learning. He employed a list consisting of pairs of words and numbers. The subject's task was to guess the number corresponding to each word as it was read to him. After each response the experimenter told him whether he was right or wrong. Since only a limited sequence of numbers was used, some hits would soon be made by chance. On subsequent presentations of the list the subject was to try to repeat the correct responses already discovered and to find more. The most conspicuous result of this experiment, as one should expect, was that correct responses were repeated from presentation to presentation to an extent far exceeding chance expectation. Incorrect responses were also found to be repeated more frequently than chance alone would lead one to expect,

[1] Wallach, H. and Henle, M., An experimental analysis of the Law of Effect, *J. exper. Psychol.*, 1941, **28**, 340–349.

and the more so the closer they were in the series to a correct response. This 'gradient' of repetition of wrong responses was cited by Thorndike as further evidence for an automatic action of reward.[2]

Yet the meaning of this experiment remains ambiguous when confronted with our question: Is it true that reward acts directly and automatically upon a stimulus-response connection or is reward effective only as a part of the larger context in which it appears? To be sure, there are in this experiment certain stimulus-response sequences which are rewarded and subsequently learned. But in addition the situation, if properly understood by the subject, is such that it is the sensible thing to remember the correct stimulus-response pairs and to reproduce the responses, because it is clear from the instructions that on successive presentations of a particular word the same response will be correct. Some will probably refuse to recognize here the existence of two separate elements, the reward of single stimulus-response sequences and the sensible nature of repetition of correct responses. They will argue that these are merely two different aspects of the same thing, the first a statement of the function, and the second a description of the way this function presents itself phenomenally. Indeed we know of no previous investigation which does not contain simultaneously these two elements.

In the experiments of our investigation, these two elements of Thorndike's experiment are separated experimentally. In our previous experiments, certain stimulus-response sequences were rewarded as in Thorndike's case, but under conditions where no sensible reason existed to reproduce or to avoid rewarded responses. Subjects were told that the same list of words would be read to them repeatedly, but that the numbers assigned to the words would vary in a random fashion from one presentation of the list to the next. Correct guesses were, however, rewarded as in Thorndike's experiment. If reward has that immediate and mechanical effect of Thorndike's hypothesis, one would expect to find rewarded responses repeated more frequently than incorrect ones, or at least to be found to be better remembered when the subject is asked to recall the responses he has made.

There arises of course one technical difficulty in this new procedure: one must manage to keep the subject seriously interested in his performance; otherwise the reward may lose its significance. It was for this reason that we disguised the experiment as a test of extra-sensory perception. Most of the subjects were interested in this problem or in their own achievement in this obscure realm of performance and watched the effects of their responses carefully.

[2] Thorndike, E. L. An experimental study of rewards. *Teach. Coll. Contrib. Educ.*, No. 580. New York: Teachers College Bureau of Publications, 1933.

In this manner we set up an experiment which duplicated Thorndike's in all essential respects except for its character of a test of extra-sensory perception and except for the failure of correct responses to continue to be correct from presentation to presentation of the list of words. The effect of reward could thus be tested under conditions where no sensible reason existed for either repetition or avoidance of correct responses.

In such experiments, reported in the previous paper, reward proved to be ineffective in bringing about repetition of responses. Rewarded responses were not repeated more frequently than wrong ones, and reward was further found not to have any differential effect upon recall of rewarded and punished responses. Furthermore, the gradient of repetition of wrong responses, which Thorndike considered to be convincing evidence of an automatic effect of reward, was absent in our experiments. Nor did we obtain the general high level of repetition of wrong responses which was conspicuous in Thorndike's experiments.

The difference between Thorndike's results and ours with respect to repetition of incorrect responses suggests that the high level of repetition of such responses and the gradient are related to the subject's knowledge of the continued correctness of correct responses and his resulting learning motive. The present experiments were designed in part to cast light upon this point. They were designed also to obtain further information about the repetition of correct responses: they were set up to determine whether, by a mere change of instructions about the continued correctness of correct responses, results could be shifted from those of our previous experiments to Thorndike's.

As in the case of our previous experiments, Experiment I was presented to the subject as a test of extra-sensory perception. Initially the subject was informed that a response which was correct on one trial would not necessarily be correct on the next. After a certain number of trials under these instructions, the subject's expectation with regard to the continued correctness of responses was changed. The subject was told that correct responses would continue to be correct from trial to trial: repetition of previous hits thereby became sensible. The experiment was still presented to the subject as a test of extra-sensory perception, however, and all other features of the procedure likewise remained constant. Thus, if the effect of reward upon repetition of response depends upon whether or not such repetition is sensible, the level of repetition of correct responses should be very different in the two halves of the experiment. If, however, the effect of reward is an automatic one, the two parts of the experiment should not yield essentially different results.

A test of recall was introduced after each part of the experiment. This was done, as in the earlier work, to reveal any possible influence of reward upon memory which, for one reason or another, might not become overt in the behavior elicited by our 'test of extra-sensory perception.'

The following instructions were read to the subject: "You have perhaps heard that experiments on extra-sensory perception are being performed by Mr. Rhine at Duke University. Subjects are asked to name items which they cannot possibly perceive in any ordinary sense of the word. If they are able to make hits more frequently than they would by chance alone, it must then be by extra-sensory perception.

"We are dealing with the same problem by a method somewhat different from Mr. Rhine's. We have prepared a list of pairs of words and numbers like this." (At this point a sample list of word-number combinations was shown to the subject.) "We shall read you the words and you guess the numbers that go with them. If extra-sensory perception exists, you should be able to do this better than chance. If it does not exist, you will merely be guessing the number that goes with each word, and you should not make a score better than chance.

"As numbers we have used the digits from 1 to 10. So you can choose among ten different reactions in each case, and with this large number you will, of course, not make many hits. But neither did Rhine's subjects. After each judgment we shall tell you whether you were right or wrong. To encourage you further, we shall pay you half a cent for each hit you make.

"Thinking about your performance during the experiment will be definitely detrimental in a function as intangible as extra-sensory perception. So try to behave naively and give your responses as quickly as you can. Do not mind if we hurry you through the experiment; it will only help you to make a good score.[3]

"We have a list of 20 words, and this list will be presented a number of times. But the numbers assigned to the words will change in a random fashion from presentation to presentation. Thus a number which is correct on the first presentation will not necessarily be correct on the next."

The procedure indicated by these instructions was followed for four complete presentations of the list of words. The list consisted of 20 words, four of which were made correct on each trial in accordance with a prearranged pattern of positions of responses which were to be called correct no matter what the subject's reaction was. This device[4] was employed to secure even spacing of correct responses[5] rather than the random distribution of hits which would have resulted had a prearranged list of correct word-number combinations been used. The positions of the correct responses were varied in an orderly manner on consecutive presentations of the series. Furthermore, the pattern of positions of correct responses was varied from subject to subject to eliminate any influence of unequally favorable positions or individual words upon the results.

After four presentations of the list, a test of recall was introduced with the instructions: "Now will you please repeat the responses you just gave while I read the list of words. I mean try to remember your last responses." The recall test was followed immediately by instructions which were designed to change the subject's expectation with respect to the continued correctness of responses. The following instructions were read: "From now on we'll always use the same list of pairs of words and numbers. A response which is correct on one presentation of the list will also be correct on all of the following presentations. See how many hits you can make now."

An additional four presentations of the list were given under these instructions, and then a second test of recall was introduced in the same manner as before. Correct responses were ininserted (regardless of what the subject's responses were) in four positions in the list of 20 words. Two of these were given on the first of these presentations, one on the second, and one on the fourth trial. Correct responses were placed so that on the trial immediately preceding the recall test they would fall at the same positions as in the corresponding trial of the first part of the experiment. To insure an equal number of correct responses on the presentation immediately preceding the recall test in the two halves of the experiment, additional hits were introduced in case any of the correct responses failed to be repeated on the fourth trial. The additional hit was in each case given in the position immediately after the original one that failed to be repeated.[6]

As in the earlier work, two experimenters conducted the experiment, the one reading the words to the subject and informing him of the correctness of his responses, the other recording the responses. Thirty-five undergraduate students were used as subjects in this experiment.

Those results of Experiment I which concern repetition of response are presented in Table I. In the first part of the experiment,

[3] Actually the tempo of our experiment was comparable to Thorndike's.

[4] First used by Muenzinger and Dove. Muenzinger, K. F., and Dove, C. C., Serial learning: I. Gradients of uniformity and variability produced by success and failure of single responses. *J. gen. Psychol.*, 1937, **16**, 403–413.

[5] It was desirable to have evenly spaced correct responses to facilitate computation of a gradient.

[6] This procedure was not followed with the first 15 subjects of this experiment.

TABLE I

REPETITION OF RESPONSES AT DIFFERENT DISTANCES FROM THE REWARDED ONES IN EXPERIMENTS I AND II

	Rewarded Responses			Wrong Responses One Place Removed			Wrong Responses Two Places Removed			Wrong Responses Three Places Removed			Σ Wrong Responses		
	No. of Responses	No. of Repetitions	% of Repetitions	No. of Responses	No. of Repetitions	% of Repetitions	No. of Responses	No. of Repetitions	% of Repetitions	No. of Responses	No. of Repetitions	% of Repetitions	No. of Responses	No. of Repetitions	% of Repetitions
No learning motive															
Experiment I, part I	416	58	13.9	791	114	14.4	783	107	13.7	—	—	—	1664	233	14.0
Experiment II, part II	420	46	11.0	799	99	12.4	791	96	12.1	—	—	—	1680	201	12.0
Total	836	104	12.4	1590	214	13.5	1574	202	12.8	—	—	—	3344	434	13.0
Learning motive															
Experiment I, part II	209	159	76.1	417	62	14.9	415	67	16.1	278	34	12.2	1891	254	13.4
Experiment II, part I	212	147	69.3	423	58	13.7	425	45	10.6	273	38	13.9	1888	233	12.3
Total	421	306	72.7	840	120	14.3	840	112	13.3	551	72	13.1	3778	487	12.9

in which the subject was told that correct responses would not necessarily be correct on subsequent trials, correct responses were repeated in 13.9 percent of the cases. They were not repeated more frequently than incorrect responses (14.0 percent) in this part of the experiment. These results thus confirm the findings of our previously reported experiments.

Strikingly different are the results of the second part of the experiment, in which the subject was led to expect rewarded responses to continue to be correct from trial to trial. Although the level of repetition of wrong responses was about the same as that in the first part (13.4 percent), the repetition of rewarded responses was increased to 76.1 percent of the cases. It is apparent that this figure is significantly higher than the percentage of repetition of wrong responses, as well as that of correct responses in the first part of the experiment.

Positions of correct responses were rotated from trial to trial in the first half of the experiment, so that there were no cases in which a correct response appeared repeatedly at the same position. Therefore the two parts of the experiment are strictly comparable only when such cases are excluded from consideration in the second part. For this reason it may be more appropriate to consider only first repetitions of correct responses in the second part of the experiment. The percentage of such repetitions (62.9 percent) is still significantly higher than that of wrong responses and of correct ones in the first half of the experiment.

The results of the present experiment again show that reward *per se* has no automatic effect upon repetition of responses. When the sensible nature of the repetition of rewarded responses is eliminated, these responses are not repeated more frequently than wrong ones. When, however, this repetition is made sensible, results are obtained of the type of those upon which Thorndike based his Law of Effect.

In our previous paper we pointed out that a motive to learn and to reproduce correct responses, rather than the single rewards, is responsible for the repetition of correct responses. The crucial difference between Thorndike's experiments and our 'extra-sensory perception' experiment consists in the fact that this learning motive is present in his experiments but does not arise under the 'extra-sensory perception' conditions. In the experiments just reported, a striking change from the one type of performance to the other is brought about merely by the shift of instructions with respect to the continued correctness of correct responses.

The results on intentional recall of responses lead to a similar conclusion. The relevant data are presented in Table II. In the first half of the experiment rewarded responses were not recalled more

frequently than wrong ones (20.7 percent against 21.8 percent). In the second part, however, correct responses were recalled significantly more often than wrong ones (66.9 percent against 23.1 percent). When we consider only those cases in which a correct response first appeared on the trial immediately before recall, the percentage of correct responses recalled (53.6 percent) is still significantly greater than that of wrong responses or of correct responses in the first half of the experiment. Whether or not correct responses are recalled more frequently than incorrect ones is found again to depend upon the nature of the instructions.

TABLE II

RECALL OF REWARDED AND WRONG RESPONSES IN EXPERIMENTS I AND II

	Rewarded Responses			Wrong Responses		
	No. of Responses	No. Recalled	% Recalled	No. of Responses	No. Recalled	% Recalled
No learning motive						
Experiment I, part I	140	29	20.7	560	122	21.8
Experiment II, part II	140	38	27.1	560	105	18.8
Total	280	67	23.9	1120	227	20.3
Learning motive						
Experiment I, part II	121	81	66.9	579	134	23.1
Experiment II, part I	140	87	62.1	560	120	21.4
Total	261	168	64.4	1139	254	22.3

In Experiment II the procedure of Experiment I was reversed. Under otherwise unchanged conditions, the subject's expectation that rewarded responses would continue to be correct from trial to trial was subsequently changed to an expectation that they would not necessarily continue to be correct.

Experiment II differed from Experiment I also in that it was not presented to the subject as a test of extra-sensory perception. The first part of Experiment II was to resemble Thorndike's procedure in every respect. It was meant as a check on the second part of Experiment I, which was given as a test of extra-sensory perception. When the instructions were changed to abolish the expectation of continued correctness of responses, the extra-sensory perception motive could not plausibly be introduced. The subjects therefore found themselves in a situation where they had to guess numbers in a senseless fashion. In our previous paper we warned against employing a psychological situation the nature of which, in the absence of a motive, would be undefined for the subject to a considerable degree. The variety of individual interpretations which may arise in such a situation might be expected to produce widely different results in

different subjects. It will be seen below that this actually happened to some extent in the second part of Experiment II.

The following instructions were first read to the subject: "We have a list of words here, and one number from 1 to 10 has been arbitrarily assigned to each word. We shall read you the words, and you try to find the numbers that go with them. After each judgment we shall tell you whether you were right or wrong. The same list will be presented a number of times, so that you will have a chance to find more and more correct numbers. Try to make as many hits as you can. To encourage you to do this, we shall pay you half a cent for each hit you make, not only for the newly found ones, but also for each time you repeat a previous hit.

"To show you how this is done, we'll try it with a sample list first." The subject was then given practice with a short sample list of words which was presented three or four times.

The procedure indicated in these instructions was followed for four presentations of the list of words, after which a test of recall was introduced as in Experiment I. The following further instructions were then read to the subject: "Now let's go on with the number-finding. We'll go on as before except that from now on the numbers assigned to the words will change in a random fashion from presentation to presentation. Thus a number will not be correct twice for the same word unless this happens by chance." Four additional presentations of the list were then given; they were followed by a second test of intentional recall. In every other respect this experiment duplicated the conditions of Experiment I. Thirty-five subjects served in Experiment II.

With respect to repetition of response, Experiment II confirms the results of Experiment I. The relevant data appear in Table I. In the first part of the experiment, in which the subject was told that rewarded responses would continue to be correct, these responses were repeated in 69.3 percent of the cases, wrong ones in 12.3 percent. The difference is statistically significant. Likewise first repetitions of correct responses occurred significantly more frequently (59.0 percent) than repetitions of wrong responses. On the other hand, in the second part of the experiment, where subjects were told that correct responses would not necessarily continue to be correct, rewarded responses were repeated no more frequently than wrong ones (11.0 percent as compared with 12.0 percent of the cases).

The results on recall are presented in Table II. As in Experiment I, that part of the experiment in which the subject was told that correct responses would continue to be correct showed a significant difference in recall of rewarded and of wrong responses (62.1 percent as against 21.4 percent of the cases). When we consider only those correct responses which were new on the trial immediately preceding the recall test, recall of rewarded responses (48.8 percent) is still significantly higher than that of wrong responses.

In the second part of Experiment II, where the subject was told that correct responses would not necessarily continue to be correct from trial to trial, the difference in recall of rewarded and wrong responses does not entirely disappear, as it did in the corresponding case in Experiment I. Here rewarded responses were recalled in 27.1 percent of the cases, wrong ones in 18.8 percent. The difference in recall of rewarded and wrong responses in this part of Experiment

II is not, however, statistically significant. It will be further noted that the records of a very few subjects are responsible for what difference does exist. Table III shows that while only one subject of Experiment I recalled as many as three rewarded responses in this part of the experiment, four subjects of Experiment II recalled that many. These subjects entirely account for this difference between the two experiments. It is likely that in the case of these subjects there was some carry-over, from the first part of the experiment, of a motive to memorize correct responses, or that some other atypical attitude developed in the psychologically little defined situation of

TABLE III

RECALL OF CORRECT RESPONSES IN PARTS OF EXPERIMENTS I AND II WITHOUT LEARNING MOTIVE

	No. of Recalls per Subject	0	1	2	3	4	Total
Experiment I, part I.....	No. of subjects	16	10	8	1	0	35
	No. of recalls	0	10	16	3	0	29
Experiment II, part II....	No. of subjects	13	12	6	2	2	35
	No. of recalls	0	12	12	6	8	38

this part of the experiment. The results of Experiment II with respect to recall would thus seem to confirm those of Experiment I to the extent to which it is possible to effect the necessary shift of set from one part of the experiment to the other.[7]

It will be noted that similar results were obtained in the second part of Experiment I and the first part of Experiment II despite the difference in instructions in the two cases. Experiment II, it will be recalled, was started with Thorndike's instructions, while the corresponding part of Experiment I (the second part) was left within the framework of a test of extra-sensory perception. It now appears that this difference is of no importance, but that the significant thing is the instruction about the continued correctness of responses. And a mere change of instructions in this respect is sufficient to shift the amount of repetition and recall of correct responses from a chance level to a level comparable to that which Thorndike found in his experiments.

While, by the use of appropriate instructions, we obtained results like Thorndike's on rewarded responses, this was not true in the case of wrong responses. The parts of our experiments in which the learn-

[7] It is interesting to note that where there was carry-over of the motive, this occurred only when recall was tested under the influence of an intention to recall, and did not appear in the case of repetition of responses in the absence of any intention to repeat them.

ing motive was present, and which were thus comparable to Thorndike's, failed to duplicate his results in two respects. Thorndike found that wrong responses were repeated approximately twice as frequently as they would have been repeated by chance, and that they were repeated the more frequently the closer they were in the series to a rewarded response. In the comparable part of our experiments, however, repetition of wrong responses remained close to a chance level. In the combined results of this part of the two experiments wrong responses were repeated in only 12.9 percent of the cases. Furthermore, while the combined results of the two experiments show slight differences in the direction of a gradient of repetition of wrong responses adjacent to a correct one, these differences are completely insignificant as compared with the differences Thorndike obtained. (*Cf.* Table I.)

In our previously reported experiment there was also a low level of repetition of wrong responses and no gradient was present. We pointed out in the previous paper that this discrepancy between our results and Thorndike's excluded his explanation of his findings, namely that the mere fact that a response has once been made tends to bring about its repetition when the occasion again arises. We argued that Thorndike's results must rather be a consequence of the motive of his subjects to learn correct responses.

The results of the appropriate parts of our present experiments now show, however, that this is not the only necessary condition for the occurrence of the high level of repetition of wrong responses and of the gradient. These parts of the present Experiments I and II differed from Thorndike's in the number of times the list of words was presented to the subject and in the number of words in the list. Thorndike used longer lists of words than we or gave more repetitions of the list, or both. It would appear that some factor associated with the extreme crowding [8] in the longer experiment, in addition to the existence of the learning motive, is a necessary condition of Thorndike's findings on repetition of wrong responses. This supposition finds some support in the results of unpublished experiments in which an attempt was made to study the gradient under simpler conditions than Thorndike had set up. To this end both the length of the series and the number of presentations were reduced. Under these conditions neither a high level of repetition of wrong responses nor a gradient was obtained.

[8] 'Crowding' is the term by which v. Restorff and others refer to an agglomeration of similar items in one series which is known to have a disturbing effect upon retention. *Cf.* H. v. Restorff, Analyse von Vorgängen im Spurenfeld. I. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychol. Forsch.*, 1933, **18**, 299–342. *Cf.* also C. E. Buxton and E. B. Newman, The forgetting of 'crowded' and 'isolated' materials. *J. exp. Psychol.*, 1940, **26**, 180–198.

Experiment III was set up to obtain further information on this point. It was introduced with the instructions used in the first part of Experiment II (in which the subject was told that correct responses would continue to be correct from trial to trial), but it duplicated our previously reported experiment in which the 'extra-sensory perception' technique was employed in the length of the list of words and in the number of repetitions of the list. A list of 20 words was used, and was presented 21 times to the subject.[9] Responses were right or wrong depending upon their correspondence with a prearranged list of correct word-number combinations. Twelve subjects participated in Experiment III.

The results of Experiment III are presented in Table IV. Wrong responses were repeated in 16.3 percent of the cases. This figure is significantly higher than the percentage of repetition of wrong responses in the earlier reported 'extra-sensory perception' experiment (10.8 percent), and it approaches the level of repetition of wrong responses which Thorndike found. The slight differences in percentage of repetition of wrong responses in different positions with respect to rewarded ones are in the direction of a gradient.

It will be observed that our results differed from Thorndike's in another important respect. While Thorndike reported repetition of rewarded responses in 'about 50 percent' of the cases, our subjects repeated rewarded responses in 86.2 percent of the cases. That this difference is probably associated with any remaining difference between our results and Thorndike's is indicated when we eliminate those subjects who showed the highest level of repetition of correct responses. When we eliminate the four subjects who repeated rewarded responses more than 90 percent of the time, our results become more like Thorndike's. In the case of the remaining eight subjects, wrong responses were repeated in 17.2 percent of the cases. Furthermore the differences in percentage of repetition of wrong responses at different distances from rewarded ones are more striking; the gradient for these subjects is of the order of magnitude of those obtained by Thorndike and by Muenzinger and Dove.

The fact that in Experiment III we obtained results like Thorndike's with respect to repetition of wrong responses, but failed to do so in the comparable parts of Experiments I and II confirms our supposition that some factor associated with the extreme crowding in the long experiment is a necessary condition of these results. Thus the high level of repetition of wrong responses and the gradient are the product of at least two unrelated conditions. Extreme crowding alone is insufficient to produce them, as shown by our earlier

[9] Muenzinger and Dove confirmed Thorndike's results in an experiment using a list of 20 words with approximately the same number of presentations as in our Experiment III.

TABLE IV

REPETITION OF RESPONSES AT DIFFERENT DISTANCES FROM THE REWARDED ONES IN EXPERIMENT III

	Rewarded Responses			Wrong Responses One Place Removed			Wrong Responses Two Places Removed			Wrong Responses Three Places Removed			Σ Wrong Responses		
	No. of Responses	No. of Repetitions	% of Repetitions	No. of Responses	No. of Repetitions	% of Repetitions	No. of Responses	No. of Repetitions	% of Repetitions	No. of Responses	No. of Repetitions	% of Repetitions	No. of Responses	No. of Repetitions	% of Repetitions
All subjects	1611	1389	86.2	1598	269	16.8	780	131	16.8	374	58	15.5	3189	521	16.3
8 subjects with repetition of correct responses below 90 percent	909	735	80.9	1086	204	18.8	562	94	16.7	292	45	15.4	2291	394	17.2

experiment with the 'extra-sensory perception' technique; here a long experimental series was employed, but a learning motive was absent. Likewise learning motivation alone is insufficient; it was present alike in Experiment III and in the second part of Experiment I and the first part of Experiment II. Yet in the two latter cases, where the amount of crowding was not excessive, the gradient was absent and the level of repetition of wrong responses was conspicuously lower than in Experiment III.

It has been shown earlier that Thorndike's results on repetition of correct responses must be reinterpreted. The present experiments now point to the necessity of a reinterpretation of his results on wrong responses as well. It would appear that the gradient, which Thorndike interpreted as a senseless spread of an equally senseless automatic effect of reward, is the product of very special conditions, the presence of a motive to learn right responses and the high degree of crowding. These same very special conditions appear to account for the general high level of repetition of wrong responses which Thorndike attributed to a tendency for repetition of responses once made.

In what way can learning motivation and crowding produce these results? In our earlier paper we distinguished two aspects of the learning motive: a motive to *learn* correct responses in order to be able to repeat them later and a tendency to *reproduce* correct responses at their proper places. That both factors were operating to produce Thorndike's results on correct responses has been shown in the previous paper. In the case of wrong responses there must also be an influence of some aspect of this learning motivation since the high level of repetition of wrong responses is only observed in its presence. Since this level is clearly above chance, there must be some basis for it in the memory of the subject. It will be noted that the level of repetition of these wrong responses in our experiments has been uniformly below the level of incidental memory of wrong responses as revealed by the recall test (about 20 percent), *i.e.*, memory of responses made in the absence of a motive to learn them. The amount of this incidental memory is thus high enough to account for the level of repetition of wrong responses. That the motive to learn responses cannot have much effect on this level is shown by the very small difference between recall of incorrect responses in experiments in which learning motivation was present and those in which it was absent (22.3 percent against 20.3 percent for Experiments I and II. See Table II). It must, therefore, be the other aspect of the learning motivation, namely the tendency to reproduce correct responses at their proper places, which somehow also affects the level of repetition of wrong responses.

It has been pointed out that the high level of repetition of wrong responses must have some basis in memory, but it is equally apparent that if the wrong items were specifically remembered *as wrong items*, such repetition would probably not occur, because the subject would be in a position to vary these responses. If, however, some of the incidentally recalled responses were remembered without the correct coördination between words and responses and the reactions of the experimenter, they would be reproduced along with the correct responses when the motive to recall these is present. Thus they would account for the level of repetition of wrong responses. It is plausible that extreme crowding will in some cases affect just that coördination, and in this way by interfering with the specific memory of wrong items as wrong, contribute to the occurrence of the high level of repetition of wrong responses.

Summary

In a previous paper we reported experiments which showed that reward as such, in the absence of a learning motive, has not the effect of favoring repetition or recall of responses. Rewarded responses in these experiments were not recalled more frequently than wrong ones. In the work reported here a change was brought about in the course of a single experiment from this type of performance to that which Thorndike found, namely high recall of correct responses. This was accomplished merely by a shift of instructions about the continued correctness of responses.

Further data on repetition of wrong responses were reported. Learning motivation and extreme crowding were found to be necessary conditions for the occurrence of the high level of repetition of wrong responses and of the gradient which Thorndike obtained.

(Manuscript received June 18, 1941)

THE RELATION OF OVERT MUSCULAR DISCHARGE TO PHYSIOLOGICAL RECOVERY FROM EXPERIMENTALLY INDUCED DISPLACEMENT

BY G. L. FREEMAN AND J. H. PATHMAN

Northwestern University

Statement of Problem

Freud, Lewin and many others have written extensively about the tendency of internally aroused excitation to seek channels of overt behavioral discharge and have developed 'psychic' energy constructs to explain how the blocking or frustration of habitual emotional outlets affects organic equilibrium. Our approach to the problem of frustration is, in contrast, more objective, and our interpretations are strictly physiological. Overt behavior is considered a means of discharging bodily energies aroused by internal tissue needs and associated external displacing stimuli. These overt discharges function, by hypothesis, to maintain the homeostatic balance of the total organism.[1] If the 'normal' and habitual outlets of aroused energies are blocked by some inhibitory process, the internal disturbance persists by aid of 'maintaining stimuli' until overt behavior of some character can effect equilibratory discharge.[2] All varieties of behavior, including skeletal, verbo-motor, and ideo-motor types may be used together or singly in the discharge of aroused excitation. In general, however, it is thought that the greater the internal excitation the greater will be the tendency of discharge through the gross movements of the skeletal musculature.

Apparent verification of the last statements appears in empirical observations. For instance, the person who is displaced or 'upset' (internally aroused) by 'unpleasant' news in a letter, finds that he calms down more quickly if he goes for a walk instead of sitting still and raging internally because no truly appropriate action can be taken. This does not mean, of course, that walking *per se* discharges

[1] For a more detailed account of the psycho-physiology of frustration behavior *cf.* G. L. Freeman (7).

[2] In our thinking, there is nothing mystical about the equilibratory response. It is, by hypothesis, any activity which discharges more excitation than it feeds back into the central nervous system. Increased muscular tension characteristically accompanies or is caused by internally developed 'emotional' excitation, which reinforces the central excitation (via backlash proprioception) rather than reduces it. An overt muscular reaction which removes the source of 'irritation' and creates a minimum of new excitation in its place, is, by hypothesis, an equilibratory response.

excitation or that it will completely reëquilibrate the organism. As a matter of fact, too much walking would probably create more internal disturbance by backlash excitation rather than reduce this state by such overt expression. Freeman (**9**) has already touched upon this problem in his critique of Hollingworth's studies on chewing as a technique of relaxation; but there is great need for further investigation of the dynamics of equilibratory response.

This paper is concerned with the relations between overt muscular discharge and recovery from the internal impact produced by experimental displacements. Our attack rests upon the assumption that the course of internal disequilibration and recovery may be observed through peripheral records of autonomically controlled activity (such as palmar skin conductance) taken periodically before, during, and after experimental displacement.

Darrow and Heath (**5**), Freeman (**6, 8**) and Krasno (**14**) have previously reported that physiological activity (blood pressure, palmar sweating, breathing) increases during startle stimulation and tends to return to the previous baseline following removal of the external causes of the arousal. The ratio of arousal to recovery in the physiological measure has been termed the *R.Q.* (recovery quotient). This quotient appeared to be a significant objective index of neuroticism and emotionality; it was suggested that unstable persons showed certain discrepancies in the ratio between the bodily energies aroused by displacing conditions and the overt reactions adequate to discharge them. The specific implication of this previous work was that low *R.Q.*'s (failure of the measures of autonomically controlled activity to show rapid recovery) were related to inhibition of overt muscular discharge. The present report constitutes a check upon this hypothesis.

In order to quantify the relation of internal arousal and the overt expression of such energies, standard situations had to be devised wherein relative amounts of overt activity could be summated and wherein there would be little occasion for excessive behavioral discharge after experimental removal of the displacing stimuli. Under such conditions, by hypothesis, arousal and discharge factors should closely parallel each other, that is, the person with rapid internal recovery should show a high degree of overt discharge.

Experimental Procedures

Palmar skin conductance[3] and skeletal movement measures were obtained on 24 male students in four types of displacing situations: startle (pistol shot), motor conflict (antagonistic finger reactions), verbal conflict (word association test involving personal histories), and dis-

[3] Actually, the original measures were in ohmic resistance. These were changed to micromho conductance units, since the latter more closely parallel the major physiological determinant of resistance, namely palmar sweating. *Cf.* Darrow (**1**).

traction (sensory discrimination during extraneous stimulation). Tests were taken in pairs on four different days, so that sets of measures were obtained for each test, once when it was given the first in a pair and once when it followed another test after a standard period of rest. Insofar as possible, tests were counterbalanced to minimize intra-serial effects.

The standardized procedure in all 4 test experiments was (*a*) allow the subject to relax on a cot until his palmar-skin conductance showed no major fluctuations [4] for five successive minutes (basal activity of rest); (*b*) introduce an 'emotionally' displacing stimulus; (*c*) remove such stimulation; (*d*) observe the rate of return in palmar-skin conductance and movement measures from the levels they had reached during the displacement. These post-test readings were taken each minute for five successive minutes [5] after cessation of the displacing stimulus.

Method of Measuring General Internal Arousal.—We used palmar-skin conductance as an index of the degree of general internal arousal produced by displacing stimulation. This measure is commended for the following reasons: (1) it is one of the most sensitive indices of increments in nervous activities obtainable in the intact organism; (2) it correlates well with the records of total oxygen consumption obtained under basal conditions; (3) it is markedly influenced by the activity of the autonomic nervous system and also by the tensions of the skeletal muscles; and (4) it shows a relatively high test-retest reliability as compared with other physiological measures of internal arousal, such as oxidative metabolism.[6]

Method of Measuring Overt Skeletal Activity.—Quantitative measurements of the whole amount of overt skeletal movement in a given situation is extremely difficult. For one thing, different *S*'s may 'express' their arousal via different muscle groups, and a measure limited exclusively to one part of the musculature will not reflect in equal degree the changes occurring in various places.

A rough index of the general restless activity of the gross musculature can be obtained in terms of movements of the cot on which the subject lies. This method has been used to advantage by many researchers, including Johnson (**12**) and Kleitman (**13**). In our adaptation of the method, an inflated air mattress was connected with a metal bellows the expansion of which[7] caused an attached needle to push a sandpaper marker along a calibrated groove. In order that movement records would be comparable from subject to subject and from day to day, a constant air pressure was kept within the pneumatic system.

It is recognized that this method did not adequately record motor discharge through the vocal apparatus, nor did it measure many other slight increments in muscular activity such as might accompany thought. Furthermore, this measure of mattress movement was differentially affected by the four types of emotionally displacing stimuli. For example, a test requiring definite manual movements affected the mattress movement more than a test requiring only verbal reactions. Nevertheless, since such differential effects presumably were relatively constant for all subjects, they did not influence individual differences in either composite movement scores of all tests or movement scores for a given test.

The Test Situation.—Since we were interested in the relations between amount of overt movement and palmar conductance changes in a variety of situations, the four tests chosen ranged from simple startle and distraction tests to more 'traumatic' situations of verbal and motor conflict. Each test will be described briefly.

Test 1. Startle Stimulus.—After *S* had relaxed to the 'basal' state, a .22 caliber blank cartridge pistol was fired without warning.

---

[4] A major fluctuation is here defined as any change in conductance greater than 1000 ohms. When we were unable to bring the individual to a stable resting level, due to his failure to relax, he was excused and returned another day. Palm-palm zinc sulphate electrodes were attached to the Behavior Resistance Box (**5**) which measures changes in standard ohmic units.

[5] Post-test readings taken at five minutes correlated highly with readings taken at ten minutes (r. 72) and in another study post-test readings at 15 min. correlated with those of total recovery time (**14**). This indicates that 5 min. readings may be used as representative of individual recovery curves.

[6] The bases for these statements are contained in a series of researches by Darrow (**2**), Darrow and Darling (**3**), Darrow and Freeman (**4**), Freeman and Giffin (**10**), and Freeman and Simpson (**11**).

[7] Only the forward expansion movement was recorded since the needle engaged the sandpaper strip when moving in one direction only.

Test 2. Motor Conflict.—This test simulated the essential conditions for the development of experimental neuroses in animals; that is, *S* was confined to a relatively restricted field of movement (lying on a cot) and the movements which would appropriately discharge the aroused excitation became impossible to make without shock punishment. *S* was instructed to push right or left keys with his index finger [8] in response to right or left stimulus lights; wrong or delayed responses were punished by electric shock delivered through the reaction keys. The series began with five trials in which either right or left light was flashed. Then followed 20 trials, of which ten were 'conflict' situations where both lights flashed at once. Finally, five more trials were given in which there were no double (conflict) stimulations. The test took seven minutes.

Test 3. Verbal Conflict.—This test attempted to arouse or reinforce existing personal conflicts to a point where they would affect changes in the general excitation level of the subject (as shown by increased palmar-skin conductance). After the subject was relaxed to the basal state the following instructions were read to assure that he would have conflicting verbal tendencies:

> "I am now going to delve into your personal life, and, since I am certain that I can discover the true nature of all your mental conflicts, I promise you that the results will be treated as confidential and will not be discussed with anyone. First, I shall read to you a series of questions concerning your behavior, and you are to think about getting ready to answer these questions a little later. It does not matter whether you decide to lie about your behavior or to tell the truth, for in either case the physiological records from our lie-detector will give us the desired index. Please ask or answer no questions now. Listen closely as I read of the various conflicts which may have entered your life, due to family, social, economic, and sex problems." [9]

The test proper was now begun with a series of stimulus words to which *S* was asked to respond with a single word. Some of these words were non-critical (such as chair), and some of them were critical (such as masturbation). The critical words were those for which the first verbal association might be blocked by an acquired inhibition or restraint.[10] The cot movements were less symptomatic of appropriate discharges in this test than in the preceding (motor conflict). A supplementary discharge measure, the *voice level* during verbal association was included.[11] The test took ten minutes.

Test 4. Sensory Discrimination.—This test was taken as representing the milder forms of annoyance. After relaxing to the basal state of rest, *S* was instructed to make a series of pitch discriminations.[12] The test began with ten easily discriminable pairs of tones, passed to discriminations of increasing difficulty, and then returned to the first ten discrimination pairs for a recheck. During the 40 interpolated pairs of tone stimuli, discriminations were made more difficult by introducing a distraction (ringing of a telephone bell intermittently). The test took seven minutes.

Results

Qualitative Features.—Wide individual differences appeared in the curves for both palmar conductance change and the accompanying overt movement. Some *S*'s who showed little conductance change to one type of displacing stimulation would be greatly aroused by

[8] Stimulus lights and reaction keys were placed at the side of the cot within easy view of *S*. Reactions could be made with a minimal disturbance of the general prone position but to a considerable extent the intensity of the finger reactions was contributory to the recorded movement changes.

[9] These statements were standard for all subjects and included such material as various words describing gratification of the sex urge, reaction to parental and economic frustration, etc.

[10] Since we are not here interested in the specific verbal responses given, these are not discussed in this report.

[11] This measure was obtained by having *S* count into a microphone and reading this 'normal' voice level from an intensity meter. From this base, increments or decrements in voice level during verbal association were computed.

[12] Record 1A of the Seashore Musical Talent Test was used.

some other kind. Some *S*'s were more restless under one type of stimulation than another. But in spite of these expected differential effects a given individual appeared to show a considerable degree of consistency in his reactions on all tests. The typical picture was one in which the initial effect of stimulation on skin conductance and movement scores was a heightened level of activity maintained without notable variation until the test was over and *S* began to relax. This suggested that the initial activity levels produced by the prolonged displacements of verbal, motor, and distraction tests could be compared directly with the activity levels produced by the momentary startle displacement. Taking as the 'basal level' in each test a point from which reaction occurred, the typical course of both measures was a pronounced increment in activity to a 'test

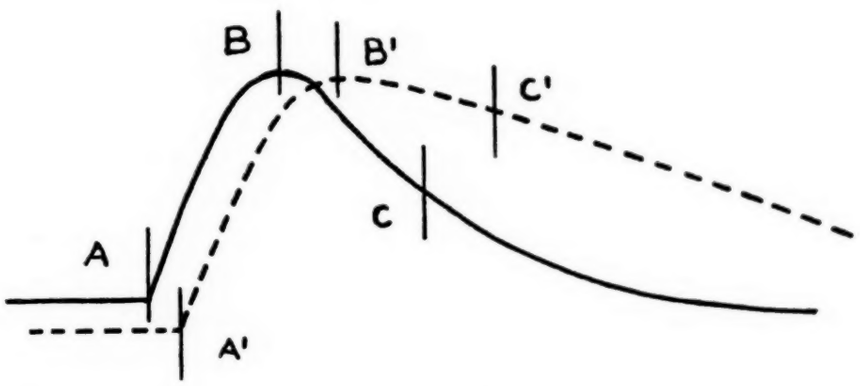

FIG. 1. Schematic diagram showing theoretical effect of displacing stimulation on palmar skin conductance and overt skeletal movement.
——— palmar skin conductance
– – – – overt activity

level' followed by a return toward the basal condition in a 'post-test level.' As seen in the diagrammatic representation of Fig. 1 this course is suggestive of a total equilibratory process (neuromuscular homeostasis), though the two response curves are not exactly parallel. The one representing internal autonomic arousal is slightly antecedent to the one representing external overt discharge. Furthermore, the recovery process of the overt movement measure is more delayed.

One of the major reasons for the failure of movement measures to decrease immediately after cessation of the displacing stimuli appeared to be an inhibition of skeletal discharge during the test situation. For example, in the motor conflict situation it was observed that many individuals 'froze'; that is, they did not make either a specific finger response or general writhing movements in spite of the fact that they received a penalty shock for not responding at once. On the other hand, there were individuals who quickly resolved the

'conflict' by striking either reaction key and taking the momentary punishment, and they tended to show rapid recovery in the conductance measures of internal disturbance. Intermediate between these reaction 'types' were those individuals who attained rather rapid internal recovery by using less specific and supplementary discharge mechanisms such as cries for help, sweating, squirming, and semi-hysterical giggling. In general, internal recovery was more rapid for 'squirmers' than for 'freezers.'

The question might be raised as to whether reaction to a standard displacement test provides a measure of the inhibition of overt discharge characteristic of the individual in a situation where he is habitually inhibited. We are inclined to believe that inhibition of overt discharge developed in the course of individual history spreads to include reaction to widely divergent situations. For example, the child who develops inhibition of overt discharge in his school life will tend to manifest this tendency in other social spheres. Since no suggestion or instruction was given to either inhibition or expression of overt reactions, in our experiments we must assume that such discharges as occurred are habitual; and the wide individual differences in the composite scores of all tests would seem to indicate that these are not an artifact of the situations.

Derivation of the Quantitative Measures.—Certain procedures for treating the raw data had to be developed before quantitative comparisons of internal arousal and overt discharge scores could be made. We need not here recount the many relationships which were tested. All we shall do is to describe the several measures which appear to have significance for the problem of this study.

Of these the most important is the *R.Q.*, a measure of the ratio of internal recovery to arousal in experimentally induced displacement. This is an integrative formula applied to the palmar-skin conductance records taken before, during, and after displacing stimulation. It is developed in the following manner:

The average conductance of the last five minutes of pre-test is treated as a base (*A* in Fig. 1) from which other changes are computed.[13] Then the percent of conductance change produced by the experimental displacement is obtained on the basis of the conductance value following the first minute of the test. The formula for this increment or arousal change (in terms of the designation of Fig. 1) is $\frac{B-A}{A}$. Next the percent conductance change following removal of stimulation is obtained by use of the conductance value at the 5th minute of the post-test period $\left($ the formula is $\frac{B-C}{A}\right)$. The complete *R.Q.* may now be computed by dividing conductance change following stimulation by conductance change caused by stimulation $\left($ in terms of the usual designations $R.Q. = \frac{B-C}{B-A}\right)$. An *R.Q.* value below 1.00 will indicate the degree (5 minutes after

[13] The values used in the formulas which follow are comparable, being in each case an average conductance value per minute of time.

displacement) that palmar-skin conductance has returned toward its previous base; whereas a value greater than 1.00 will indicate the degree to which that recovery has advanced beyond the base.[14]

In line with the procedures established in obtaining the arousal measures, movement discharge was scored in terms of the percent increment during test stimulation (in terms of the designation of Fig. C, $\frac{B' - A'}{A'}$). A similar type of *R.Q.* $\left(\frac{B' - C'}{B' - A'}\right)$ was computed for overt movement, but was found of little significance in later comparisons. This is apparently due to the fact that inthe 5-minute post-test period, movements were still being utilized to discharge excitation aroused by the original stimuli. Because of the latency in movement recovery, a longer post-test period would have been necessary to observe the return of movement discharge to its original level. Thus test and post-test scores were here combined into a measure of total movement increment $\left(\frac{(B' - C') - 2A}{2A}\right)$.

Reliability of the Measures.—None of the measures of internal arousal and external discharge can be more reliable than are the scores on which it is based. It is necessary, therefore, to determine the stability of conductance and movement data for different S's in the four test situations. We have already described precautionary procedures which would limit chances of variation in basal conductance, and Table I indicates that these precautions were effective. Not

TABLE I

TEST-RETEST RELIABILITIES * OF SKIN CONDUCTANCE AND SKELETAL MOVEMENT MEASURES

Situations	Basal Conductance	Conductance *R.Q.*	Movement Basal	Movement Incr.
Startle (*S*)	.58	.85	.50	.89
Motor Conflict (*MC*)	.63	.90	.66	.87
Verbal Conflict (*VC*)	.69	.93	.65	.89
Distraction (*D*)	.87	.85	.37	.86

* $N = 24$. Comparisons are made between first and second days of testing.

only do the basal values show a high test-retest reliability, but the measures of conductance *R.Q.* built thereon are also reliable. The basal movement scores are about as reliable as the basal conductances, but the movement increment values show greater inter- and intra-test variation for the individual, due presumptively to the failure of the air mattress to represent all the overt discharges operative in expressing internal excitation. Attention should be called to the fact that the measures were most unreliable on the distraction test.

[14] If the basal conductance (*A*) was abnormally high (due to undischarged tensions of previous stimulation?) an experimental displacement might serve as a trigger to discharge more excitation that it arouses of itself, and the post-test conductance at point *C* would fall below that of *A*, giving an *R.Q.* greater than 1. Conversely, if the conductance level at post-test (*C*) is above both basal (*A*) and test (*B*) levels, a minus *R.Q.* would be obtained. A few instances of these aberrations were obtained on this study, but the general picture is as indicated above, *i.e.*, *R.Q.* of less than 1 at the 5-minute post-test level.

Analysis of Effects of Different Stimulus Displacements on the Measures.—The relative stability of the basal measures and the effects of different types of stimulus displacements upon conductance and movement increments are shown in Table II. The average basal conductance appears more stable from test to test than does the average movement basal. The motor conflict test, as the most 'traumatic,' involved both the greatest arousal (conductance increments) and the greatest discharge (movement increments). The distraction test was

TABLE II

EFFECTS OF DIFFERENT STIMULUS DISPLACEMENTS ON CONDUCTANCE AND MOVEMENT MEASURES

Average Measures	Stimulus Situations			
	Startle	Motor Conflict	Verbal Conflict	Distraction
Cond. Basal (A) *	18.05	18.72	18.98	18.53
Peak Cond. (B)	33.17	38.54	32.38	31.29
Cond. Increment $\left(\frac{B-A}{A}\right)$	1.17	1.56	1.00	.92
Cond. $R.Q.$ $\left(\frac{B-C}{B-A}\right)$	.55	.40	.55	.48
Movement Basal (A')	3.58	4.63	2.7	2.38
Test Movement Increment $\left(\frac{B'-A'}{A'}\right)$	30.07	150.5	10.1	8.7
Test and Post-test Movement Increment $\left(\frac{(B'+C')-2A}{2A}\right)$	19.5	72.09	7.36	5.6

* Letters in parentheses indicate formula derived with reference to Fig. 1.

the least displacing. Conductance *R.Q.* is not particularly high in any tests but is least so for the motor conflict test. The *R.Q.* for the distraction test is not typical; the group as a whole showed good recovery and the low average *R.Q.* of .48 is due to two extreme deviates who were tremendously displaced by this test.

The Communality of the Measures.—The existence of common determinants of conductance and movement scores, independent of the type of displacing stimulation, is suggested by the inter-test correlations of Table III. The correlations are for comparisons between averages of two tests in the four situations studied. They indicate that to some extent a given individual tends to behave similarly in all types of displacement. More specific evidence for communality is provided by a factor analysis now in press.

Relation between Conductance (R.Q.) and Movement (M.I.) Scores.—The first test of our hypothesis that rapid recovery of internal equilibrium is a function of overt muscular discharge came from a corre-

lation (.53) of the average *R.Q.* and *M.I.* scores for all trials and tests ($t = 2.9$, $P = < .01$). Subsequently it appeared desirable to break this average relation down into those holding (1) for the four different types of displacement, (2) for such practice and expectancy factors as might have operated between first and second tests, and (3) for such superposition factors as might have influenced the second of

TABLE III

INTER-CORRELATIONS * OF CONDUCTANCE *R.Q.* AND MOVEMENT INCREMENT MEASURES

Tests	Startle		Motor		Verbal	
	R.Q.	*M.I.*	*R.Q.*	*M.I.*	*R.Q.*	*M.I.*
Startle.....						
Motor.....	.50	.62				
Verbal.....	.55	.82	.55	.66		
Distraction.....	.60	.78	.57	.78	.56	.72

* According to Fisher's test of significance the minimum value of an r at the 1 percent and 5 percent levels when N is 24 is .515 and .404 respectively.

the pair of tests given the same day. These comparisons are shown in Table IV. The greatest correspondence between degree of internal recovery and amount of external expression appears, naturally, in the motor conflict test, where the recorded overt discharge is most expressive of movements appropriate to the displacing situation. Correspondence on the startle test is less exact than expected, due possibly to the crudity of the movement indicator; that is, a momentary increment in movement due to startle may not have been sufficient to raise the test level (*B*) significantly above the basal level (*A*). The relation between recovery and movement discharge holds also for the verbal test, although here cot movement due to wiggling etc. is a less appropriate discharge than in startle and motor conflict tests. When voice level increment (a more appropriate discharge measure) is substituted for the cot movement increment in the verbal test, the correlation between arousal and discharge is .65. The distraction test shows to disadvantage in all comparisons due to its failure to constitute a sufficiently displacing situation.

The effect of practice and expectancy factors is shown by comparing the size of *R.Q.*–*M.I.* correlations for each test on the first and second day it was given. *R.Q.*'s to startle and distraction showed the highest correlation with *M.I.* scores the first time they were given, whereas *R.Q.*'s to the more traumatic motor and verbal conflicts were more positively influenced by motor discharge on the second day than on the first.

The effect of an antecedent displacement did not much alter the *R.Q.–M.I.* correlations obtained when the test was second in a pair, except perhaps for the motor conflict test. Few of the detailed correlations in Table IV approach the 1 percent level of significance.

TABLE IV

RELATIONS BETWEEN CONDUCTANCE *R.Q.* AND MOVEMENT INCREMENT (*M.I.*) MEASURES OBTAINED FOR A VARIETY OF SITUATIONS

Displacement	All Tests	1st Test	2nd Test	1st in Pair	2nd in Pair
Startle	.31	.38	.22	.29	.22
Motor	.56	.23	.59	.53	.22
Verbal	.36	.23	.41	.30	.36
Distraction	.11	.11	.05	.12	−.07

In general, the more detailed the breakdown of the *R.Q.–M.I.* correlation, the less its significance. The counterbalancing of such factors as expectancy, superposition and presentation order apparently accounts for the more significant correlations based on average scores.

Conductance R.Q. in Relation to Post-test Movement Scores.—The comparisons made in the paragraph above have been upon the basis of movement increments made in the test situation only. Post-test movements have not been taken into account. If a general drop in movement scores had occurred from test to post-test, indicating thereby that the peak discharge level had been passed, we might have successfully correlated internal (arousal) and external (discharge) *R.Q.*'s, as was the original plan. But since the majority of subjects were still moving about on the cot as much at the time of "post-test" reading as at 'test' reading we decided to treat the post-test movements as a part of test discharge. That this is legitimate is shown by the close correlation of test and post-test movement scores (r .77; t 5, p .01).[15] When average test and post-test movement increments are added together and correlated with conductance *R.Q.*, the results do not alter significantly the picture obtained by use of only the movement increment score of the test period. The correlation of average *R.Q.* with the combined test and post test *M.I.* scores is .41. While the correlation is increased to .62 for the motor conflict test and to .42 for the verbal conflict, it is reduced to .07 for the startle test—the one displacement where post-test movement was not normally a continuation of the test discharge.

[15] The breakdown of this relation according to possibly influencing factors offers higher rather than lower correlations for the *S*, *M*, *V*, and *D* situations correlations were .75, .99, .83, .82 for 1st pair and .98, .99, .77, .94 for 2nd pair in the series. First test correlations were .89, .92, .65, .86 and second test correlations were .66, .99, .82, .92. The smaller correlations for startle (*S*) and verbal (*V*) tests are especially interesting.

Pre-test (Basal) Movement as an Index of already Existing Internal Disequilibration.—The tendency for the discharge of an aroused excitation to occur gradually through general restless movements suggests that the movements of pre-test or 'basal' rest may be symptomatic of undischarged excitations from previous stimulation. Involuntary nervous movements are among the recognized symptoms of the neurotic and many writers have spoken of 'unresolved conflicts' leading to motor automatisms, restless behavior and the like. Two tests of this hypothesis are provided by our data.

The first check was a comparison of the conductance *R.Q.* of the first of paired tests with the basal movement score preceding the second test. It was found that the lower the *R.Q.* of first tests, the greater the amount of movement (over original basal level) displayed in the rest period preceding the second displacement test ($r - .42$).

The second and more general check of the hypothesis was provided by the negative correlation between the average amount of 'basal' movement with the average of conductance *R.Q.*'s for all tests ($r - .50$). These negative correlations not only appear to confirm the hypothesis suggested above, but also suggest the more general hypothesis that those persons who discharge aroused excitation most immediately and directly will recover equilibrium of internal arousal most quickly. Thus individuals who moved the most immediately after our experimental displacements were least restless in the pre-test or basal condition. One may not go too far with the use of sustained restless movements as symptoms of internal disequilibration, for even these may be blocked, and thus a 'neurotic' may escape notice because he is almost completely 'frozen' in regard to skeletal discharge.

Relation of Internal Arousal to Movement Increment.—Another significant comparison [16] is the extent to which degree of internal arousal (percent conductance increment) is paralleled by degree of overt discharge (percent movement increment). Our data are appropriate to this question only in cases where the recorded movement is fairly expressive of all the responses which are serving to discharge the aroused excitation. This of course, is not even approximately the case for many of the displacements. The aroused excitations for the verbal conflict, for instance, are most readily expressed by verbal or ideo-motor discharges, and here the restless movements made on the cot will serve mainly as incidental or 'overflow' discharge mechanisms. No one measure is likely to indicate total overt discharge with the degree of effectiveness that skin conductance changes indicate

[16] Among the many non-significant relationships for which the data were analyzed typical was the *r* of .05 between basal conductance level and movement *R.Q.* and the *r* of .02 between peak conductance level and movement increment.

internal arousal. None the less the correlations of our two measures, the one of internal arousal and the other of external discharge, is suggestive for further work. Correlation between conductance and movement increment (for average on all tests) is .19. For the one (motor-conflict) test where the movement score is probably most representative of total discharge, r is .51.

Discussion

While our results are to be regarded as tentative and exploratory in character, they appear to have important implications for behavior theory and practice. Freud, Lewin, and other dynamic psychologists have long maintained that, when an individual is aroused, repression or blocking of the appropriate channel of reëquilibratory response produces internal 'tension' which tends to persist until relieved. And from another point of view, Sherrington [17] has suggested the more direct the motor release of aroused excitation, the faster is the recovery of internal equilibrium. The skeletal muscles offer many mechanisms for discharging aroused internal excitation. However, the fact that this class of effectors is most directly under cortical control tends, in many 'inhibited' individuals, to deflect the aroused excitation to internal organ systems (such as the stomach) whose tensions continue to maintain themselves. Because of the persistent nature of these 'maintaining' (as opposed to discharging activities) total bodily equilibrium is disturbed over longer periods of time. In fact, the extremely inhibited individual may appear outwardly calm, both during and after an emotionally exciting situation when actually his behavior might better be classed as 'frozen.' Because of inadequate discharge, latent effects of the maintained disturbance may appear subsequently in such disguised outlets as tics, nervous movements, habit spasms, compulsions, tachycardia, and other psychosomatic disorders. An individual such as mentioned above, in our tests might actually express more movement during the resting state than during displacement. The more stable individual, how-

[17] C. S. Sherrington (**16**) writes as follows: "Speech, since it provides some degree of externalization of energy, may be regarded as standing between musculo-skeletal behavior (*i.e.*, action) and thought, and is also a manifestation of partial motor inhibitions. The degree of availability of these modes of cortical expression to the instinctual levels is in direct proportion to the degree of motor inhibition, because of diminishing external risk, yet the degree of relief of instinctual tension depends on the degree of sheer motor component in the expression. Thus action gives the greatest relief, thought or fantasy the least. But instinctual action carries with it the gravest external threat and thought the least grave. Speech stands midway between them with regard to both considerations, and is thus a singularly happy medium of expression. The symbols, the fantasies, with which the patient occupies himself represent a constant effort to translate the physiologic energies of instinct into a form adequate for cortico-spinal expression, or at least into the dissipation of pure thought."

ever, would rest quietly prior to displacement because of his relative freedom from residual tensions of previous situations. Our results appear to confirm essentially these hypothetical relationships. If more adequate methods were available for quantifying the total amount of overt expression, its relationships with rate of internal physiological recovery would undoubtedly be more pronounced. It must be remembered that the individual has other means of discharging aroused excitation than those movements measured in this study. Casual observation showed that some subjects who moved relatively little on the air mattress during and after displacing stimulation used primarily verbo-motor mechanisms of discharge, such as swearing at the experimenter, giggling, laughing, and talking. We had no index of the ideo-motor discharge of 'phantasy.' This may have been relatively pronounced in some subjects, particularly those who were least responsive in their overt skeletal discharges. We do not know how effective the various discharge mechanisms are for a given individual and whether these vary independently for all subjects and with all tests. Nor have we yet any certain means of knowing whether a given discharge mechanism puts back more excitation than it takes out of the central nervous system. Many further studies are here suggested; there are, presumptively, significant but unknown relationships between conductance *R.Q.* scores as measured in this study and the *type* as well as the *degree* of equilibratory discharge.

Summary

1. Emotionally displacing stimuli were given to each of twenty-four male students when they were relaxing on a cot. Degree of 'energy' mobilized to meet the displacing stimulation, and rate of return to the pre-stimulus level of reactivity were measured in terms of palmar-skin conductance changes. Amount of overt skeletal movement was measured pneumatically from an air mattress on which the subject lay.

2. Four types of displacing stimuli were employed, including a pistol shot, motor-conflict (produced by an unsolvable reaction situation), verbal conflict (word associated with personal problems) and distraction (pitch discrimination during intermittent bell ringing). Test-retest reliabilities ranged from +.85 to +.93 for *R.Q.* measures (percent recovery in palmar-skin conductance five minutes after cessation of the stimulation), and from +.86 to +.89 for *M.I.* scores (increment in overt movement during stimulation). The intercorrelations between the *R.Q.* scores in the four types of displacement ranged from .50 to .78; and the intercorrelations between the *M.I.* scores ranged from .62 to .82.

3. Correlation between *R.Q.* scores and average *M.I.* scores (four types of displacements combined) was .53, and the correlation between average *R.Q.* scores and average movement (*M*) scores during the pre-stimulus period of rest was −.50. The correlation between *M* scores during stimulation and those following stimulation was .77, and the correlation between average increment in internal arousal (percent conductance) and average overt discharge (percent movement increment) was .19.

4. The results provide tentative confirmation of the hypothesis that, within the limits of these tests, individuals tend to recover internal equilibrium most rapidly who readily discharge aroused excitation by overt muscular action, even though this is apparently unadaptive.

(Manuscript received June 24, 1941)

References

1. Darrow, C. W., The equation of the galvanic skin resistance curve: 1. The dynamics of reaction in relation to excitation background, *J. gen. Psych.*, 1937, **16**, 285.
2. Darrow, C. W., Neural mechanisms controlling the palmar skin resistance and palmar sweating, *Arch. Neurol. & Psychiat.*, 1937, **37**, 641–653.
3. Darrow, C. W., and Darling, R., Determining activity of the autonomic nervous system from measures of autonomic change, *J. Psych.*, 1938, **5**, 85–89.
4. Darrow, C. W., and Freeman, G. L., Insensible perspiration and the galvanic skin reflex, *Amer. J. Psych.*, 1935, **48**, 55–63.
5. Darrow, C. W., and Heath, L. L., Reaction tendencies relating to personality, Lashley, K. S. (Ed.), *Studies in the Dynamics of Behavior*, Chicago: Univ. of Chicago Press, 1932, pp. 332.
6. Freeman, G. L., The effect of inhibited micturition on interrupted and completed acts of unrelated origin, *J. gen. Psych.*, 1938, **19**, 277–287.
7. Freeman, G. L., Toward a psychiatric plimsoll mark: Physiological recovery quotients in experimentally induced frustration, *J. Psych.*, 1939, **8**, 247–252.
8. Freeman, G. L., A method of inducing frustration in human subjects and its influence upon palmar skin resistance, *Amer. J. Psych.*, 1940, **53**, 117–121.
9. Freeman, G. L., Dr. Hollingworth on chewing as a technique of relaxation, *Psych. Rev.*, 1940, **47**, 491–493.
10. Freeman, G. L., and Giffin, L. L., The measurement of general reactivity under basal conditions, *J. gen. Psych.*, 1939, **21**, 63–72.
11. Freeman, G. L., and Simpson, R. M., The effect of experimentally induced tension upon palmar skin resistance, *J. gen. Psych.*, 1938, **18**, 319–326.
12. Johnson, H. M., *et al*, Sleep, *Psych. Bull.*, 1926, **23**, 482–502; 1930, **27**, 1–37.
13. Kleitman, N., *et al*, *Sleep Characteristics*, Chicago: Univ. of Chicago Press, 1937, pp. 86.
14. Krasno, L. R., The effect of displacing stimuli on the subsequent respiratory activity of rats and guinea pigs. Unpublished doctoral dissertation, Northwestern Univ. Library, 1939.
15. Lindquist, E. G., *Statistical Analysis in Educational Research*, Cambridge: Riverside Press, 1940, pp. 266.
16. Sherrington, C. S., The Brain and its Mechanism, The Rede lecture delivered before the University of Cambridge, Dec. 5, 1933, Cambridge: Macmillan, 35 pp.

THE INFLUENCE OF VARIABLE TIME INTERVALS ON RETENTION OF MEANINGFUL MATERIAL

BY FRANKLIN O. SMITH †

Montana State University

It has been said many times that retention is not comparable to a photographic plate; that material recalled is not identical with material learned; that retention is not an inert, static condition, but rather dynamic in character, an organization of the original impressions in which the past participates in the present. In memory the individual relives his past experience, but in doing so he creates a new experience at least in part.

However, the new experience is not wholly new. Part of the recalled experiences may be said to be photographic reproductions of the original impressions.

The purpose of the present study was to determine to what extent the ideas in meaningful material after one or two readings which are retained and recalled are identical reproductions of the original material and to what extent and in what respect they are modified, and further to determine the influence of various time intervals or rest periods on the retention of such meaningful material.

Three types of material were selected and presented to groups of 20 to 40 subjects, students in General Psychology. The first consisted of two paragraphs, one on nitrogen taken from Popular Research Narratives and containing for the purpose of this study 28 units of thought, single words or phrases; another paragraph on the Philippines from Scheideman's Lecture Demonstrations for General Psychology containing for our purpose 58 units of thought. The second type was the following sentence adapted from Lewes, Problems of Life and Mind. George Henry Lewes wrote: "Science is penetrating everywhere and slowly changing men's conception of the world and of man's destiny." This sentence contains, again for our purpose, 12 units of thought. The third type of material was a list of 20 related words selected from a description of a now famous football game.

In the first experiment a group of 42 subjects read the paragraph on Nitrogen and reproduced it immediately. After rest periods of 1 minute, 2 minutes and 4 minutes respectively, in which the subjects did nothing, the paragraph was reproduced without rereading. The paragraph was then reread and reproduced immediately. A control

group of twenty subjects read the paragraph twice and reproduced it immediately.

The reproductions were scored so as to show the number of ideas correctly reproduced, the number essentially correct but with some changes in verbal form, the number mainly correct but modified somewhat in content, the number wholly wrong, the number correct or partly correct but transposed, and the number of ideas in the reproduction which were not in the original. The following symbols were chosen to represent the various kinds of reproduction: *C* = correct both in content and in verbal form. *V* = verbal form changed although correct in content. *M* = content as well as verbal form modified. T = transposed. *W* = wrong and *N* = a new idea which was not in the original paragraph.

TABLE I

Paragraph (Nitrogen) read once, reproduced immediately, and without rereading reproduced after rest periods of 1 minute, 2 minutes, and 4 minutes, respectively; then reread and reproduced immediately. Column 7 contains the results of the control group and column 8 the difference between the experimental and control groups.

1	2 0	3 1 m	4 2 m	5 4 m	6 Reread	7 Control	8 Diff.
C	3.1	3.5	3.7	3.9	7.5	4.5	3.6
V	4.4	6.4	7.3	4.9	7.5	5.8	1.7
N	.7	.9	1.0	.8	1.1	.6	.5
T	.3	.3	.4	.4	.2	.1	.1
W	.2	.3	.3	.3	.0	.0	.0
N	.3	.2	.2	.2	.0	.0	.0

The results of this experiment are presented in Table I which tells its own story. The average number of correct reproductions is nearly constant for no interval, 1 *m*, 2 *m* and 4 *m* intervals, an increase of .8 from 0 interval to 4 min., but an increase from 3.9 after 4 min. rest to 7.5 after rereading. The control group with two readings and no rest period had an average of 4.5 correct. The difference between two readings with a 4 min. rest period between and two readings with no rest period is 3.

The only other item of special interest in this table is the number of ideas correctly reproduced but with slight verbal changes. The increase from 4.4 with no interval to 6.4 with 1 min. rest and 7.3 with two min. rest is not in agreement with the other results in this table. The difference of 1.7 between two readings with a 4 min. rest between and two readings with no interval is significant. The results appear to show that whereas there appears to be only slight improvement after a single reading for different rest periods, there is considerable improvement from two readings with a rest period be-

tween as compared with two readings with no rest period. The other results shown in this table are only of slight significance.

One further remark about these results. The averages on the whole show only slight improvement for different rest intervals. Individuals, however, vary quite widely in the amount of improvement. In one case the number of correct reproductions changed from 4 with no interval to 15 with four-minute interval. In another case the change was from 7 to 18. At the other extreme the change was downward from 5 to 2.

The second experiment included two experimental groups and a control group. The first group read the paragraph (the Philippines) once and after 8 minutes interval reread the paragraph and repro-

TABLE II

Paragraph (Philippines) read once and after 8 minutes reread and reproduced by one group; read once and reread and reproduced after 16 minutes by second group. Column 4 shows the control group, column 5 the difference between 8-minute interval and control group, and column 6 the difference between 16-minute interval and control group.

	8 min.	16 min.	Control	Difference	
C	15.8	18.6	11.6	4.2	7.0
V	18.8	18.5	15.6	3.2	2.9
M	2.9	3.0	2.8	.1	.2

duced it. The second group read the paragraph once, rested 16 minutes, then reread and reproduced. The control group read the selection twice and reproduced it at once. The results are presented in Table II which again tells its own story. The greatest gain is in the number of correct reproductions although the actual number of correct in content but with changes in verbal form is larger than number strictly correct.

The number of wrong, transposed and new ideas, although some of these were present, was not significant for the group.

The third experiment was similar to the first except that the material was a single sentence referred to previously, and the longest interval was 8 minutes instead of 4. Table III contains the results.

TABLE III

A sentence of twenty words was read once, reproduced immediately, and without rereading reproduced after periods of 1 minute, 2 minutes, 4 minutes and 8 minutes, respectively; then reread and reproduced immediately. Columns 8 and 9 same as 7 and 8 in Table I.

1	2 0	3 1 m	4 2 m	5 4 m	6 8 m	7 Reread	8 Control	9 Diff.
C	6.6	6.0	6.1	6.0	6.1	8.6	7.4	1.2
V	1.2	1.1	1.2	1.3	1.3	1.4	1.3	.1
M	1.0	1.2	1.3	1.3	1.3	.8	1.2	.4

With the sentence 50 percent of the material was correctly reproduced after a single reading for all intervals, and about 70 percent after two readings with an 8 minute interval between the readings. In the case of the short paragraph (Table II) only 12 percent was correctly reproduced and 27 percent with the second reading after 8 minutes. The fourth experiment is concerned with a list of 20 related words as before described. The procedure was exactly the same as in experiment one. The results appear in Table IV. Slightly more

TABLE IV

A List of 20 Words Read Once and Reproduced after Rest Periods as in Table I

1	2 0	3 1 m	4 2 m	5 4 m	6 Reread	7 Control	8 Diff.
C	10.8	10.5	10.5	10.9	16.0	11.0	5.0
V	.5	.6	.6	.5	1.4	0	1.4
T	1.7	2.2	2.4	2.4	2.8	2.3	.5

than 50 percent of the words were reproduced for all intervals and 80 after the second reading following an interval of 4 min. and a gain of 44 percent over the control group. The number of words correctly reproduced but with some change in verbal form is, as might be expected, relatively small while the number of transpositions is larger than in any other type of material. Rest periods show the greatest influence on transpositions in the case of a list of words.

It appears from these results that the influence of time intervals is partly a function of the character of the material presented. Long selections yield relatively fewer correct reproductions than short selections of about equally difficult subject matter, while a single sentence composed of relatively abstract ideas although yielding a greater number of correct reproductions after a single reading is influenced to a less degree by the rest period than the more concrete matter. The number of reproductions with various modifications as represented in the tables by the symbols *V*, *M*, *T*, *W*, and *N* also varies with the kind of material employed.

The marked increase in the number of correct and modified reproductions following a rest period over the number without a rest period is another matter. If, as Guilford, Bartlett and others suggest, learning is reorganization it may be supposed that the subjects brought to a given task (the Philippines, nitrogen or what not) an organization of response with reference to that task, that during the first reading re-organizations in which the new content was incorporated into the material of previous organization, and that during the rest period the subject was not passive and inert, but that active

organization was going on either intentionally or involuntarily in sudden and repeated flashes of the subject matter previously read. In the process of reorganization the content would undergo modification in various ways and to a greater or less degree. Transposition would seem to indicate a certain logical readjustment. A wholly wrong reproduction probably indicates the substitution of an idea previously acquired for one that is oblivescent. A new content not contained in the original may perhaps be explained by the law of similarity or contrast. The results show that the time interval has little effect on transpositions, wrong or new responses.

(Manuscript received June 16, 1941)

A NEW TECHNIC IN STUDYING THE EFFECTS OF PRACTICE UPON INDIVIDUAL DIFFERENCES

BY W. A. OWENS, JR.

Iowa State College

Studies of this problem, to the present time, have tended to follow one of two modes of attack. Either they have determined the effects of practice upon the average of the intercorrelations of a test battery, or they have investigated the effect of practice upon the standard deviation of a distribution of scores in some function or functions.

R. A. Fisher's (**3**) analysis of variance is a statistical technique possessing certain peculiar advantages in the context of this problem. For example, product-moment correlation utilizes absolute deviate ranks; it does not make available any information as to a variation within these ranks as long as they do not actually shift. The question as to whether such ranks are condensed or extended with practice is, of course, relevant here. In one form of the analysis of variance, on the other hand, an independent estimate of the contribution of individual differences to the total variation is made directly available, and this estimate can be so based as to take account of all shifts in an individual's 'relative' position.

In estimating the effect of practice upon a standard deviation representative of individual differences, the assumption is that other factors remain constant. If, for example, the test or tests employed became increasingly unreliable with practice, the standard deviation could increase in size quite independent of any change in the magnitude of individual differences. If the analysis of variance were similarly employed, the estimate of error would tend to absorb any possible shifts in the reliability of the tests as practice progressed.

The problem of establishing a zero point and equal units of measurement on the scale employed has been a long standing issue in this field. Since the analysis of variance deals with differences, or deviations, and not with absolute magnitudes, the problem of setting a zero point is obviated. Anastasi (**1**) has theoretically justified the use of standard scores from scaled distributions as equivalent indices of changes in variability with practice. In order to secure equal units of measurement, then, it is only necessary to establish the normality of all distributions used, and to score them all on a comparable basis.

An Illustrative Case

The Problem.—To obtain an estimate of the effects of practice upon individual differences in some tests of motor skills.

The Method.—Statistically, this experiment was designed for the application of the analysis of variance with two criteria of classification (**4**) individuals and administrations. Table I furnishes a schematic illustration of the technic.

TABLE I

Effect of Practice on Individual Differences

Individuals	Administrations					
	II	III	IV	VI	VII	VIII
A	425	448	425	486	511	481
B	502	469	530	531	602	586
C	514	541	648	702	730	742
Etc.	—	—	—	—	—	—

Factor	°'s Freedom	Mean Square	°'s Freedom	Mean Square
Total V.	44	——	44	——
Ind. Diff.'s	14	**10,456.37*	14	**20,075.52*
Admin. D's.	2	22,333.76	2	179.82
Error	28	1,411.18	28	842.06
	* F = 13.79 and P <.01		* F = 23.84 and P <.01	

The plan was simply to determine the magnitude of variation to be assigned to individual differences on the second, third and fourth administrations of several tests, as compared to the magnitude to be so assigned on the sixth, seventh, and eight administrations. Extraneous sources of variation were incorporated in the differences between administrations, and in the estimate of error. This latter factor would include differential rates of improvement by the various subjects, and inaccuracies in measurement. It was, thus, possible to obtain a direct estimate and comparison of the initial and final magnitudes of individual differences.

It may be objected that dichotomizing the administrations in this fashion implies that two distinct populations are present. However, this assumption may be justified, since the original sample does undergo a statistically significant change, with practice, from administrations two through four to administrations six through eight.

The subjects for the present experiment were fifteen Junior High School boys matched for age, intelligence and race, and otherwise unselected, to the writer's knowledge. Each subject received eight administrations of seven tests of motor skills (**5**). The tests were

administered in systematically varied order on a schedule calling for about eight hours of each subject's time.

The normality of distribution necessary to insure equal units of measurement was secured in the following manner. Norms for a comparable group of over two hundred subjects were available (**5**). These large norm distributions were tested, and normalized where necessary. Scores of the present experimental group were interpreted as of these larger 'master' distributions. Comparability from one test to another was achieved through equating the means and standard deviations of the master distributions. All scores were, thus, similar to the scores which McCall derives from his 'T scale.'

TABLE II

PRACTICE AND INDIVIDUAL DIFFERENCES—SUMMARY

Factor	°'s Freedom	Sum of Squares	%	Mean Square	'F'	P
Administrations 2–4						
Total V.	264	2,105,547.88		——		
Ind. Diff.'s. . . .	84	1,718,615.20	82	20,459.70	13.79	.01
Admin. D.'s. . . .	12	137,593.33		——		
Error.	168	249,339.35		1,484.16		
Administrations 6–8						
Total V.	264	2,229,093.06		——		
Ind. Diff.'s. . . .	84	1,888,625.04	85	22,583.63	11.90	.01
Admin. D.'s. . . .	12	22,941.17		——		
Error.	168	317,526.85		1,890.04		

$$t = \frac{\text{I.D.}_2 - \text{I.D.}_1}{\sigma_D} = 0.31 \text{ and } P > .50$$

The sums of squares and degrees of freedom from the parallel analyses of each test were totaled to yield two sets of summary statistics. One set was for administrations two through four of all tests, and one was for administrations six through eight of all tests. These yielded the desired estimates of the initial (2–4) and final (6–8) magnitudes of individual differences. The sums of squares and degrees of freedom were so totaled on the assumption that all deviations were about a common grand mean.

The Results.—Table II gives the results. It shows a tendency for individual differences to increase slightly, but statistically insignificantly, with practice. This result is in essential agreement with those of such earlier investigators as Thorndike (**6**), Ewert (**2**), Anastasi (**1**), and others. If this experiment should prove to be representative with respect to the effects of practice upon individual

differences in motor skills, the importance of the initial selection program in industry would surely be reëmphasized.

The Conclusion.—Individual differences in this population, and with respect to these functions, tended to remain constant with practice.

(Manuscript received June 13, 1941)

References

1. Anastasi, A., Practice and variability, *Psychol. Monog.*, 1933–34, **45**, No. 5.
2. Ewert, H., The effect of practice on individual differences when studied with measurements weighted for difficulty, *J. gen. Psychol.*, 1934, **10**, 249–285.
3. Fisher, R. A., *The Design of Experiments*, London: Oliver & Boyd, 1937, 57–58.
4. Goulden, C. H., *Methods of Statistical Analysis*, New York: John Wiley & Sons, 1939.
5. Paterson, D. G., Elliott, R. M., *et al.*, *Minnesota Mechanical Ability Tests*, Minneapolis: University of Minnesota Press, 1930, 46–63.
6. Thorndike, E. L., The effect of practice in the case of a purely intellectual function, *Amer. J. Psychol.*, 1908, **19**, 374–384.